# MODULATION OF MEMORY INTEGRATION IN THE HUMAN BRAIN

## EXPERIMENTAL INVESTIGATIONS OF MEMORY ENCODING & CONSOLIDATION

RUUD M.W.J. BERKERS

DONDERS SERIES

# Modulation of Memory Integration in the Human Brain

## Experimental Investigations of Memory Encoding and Consolidation

*Ruud M.W.J. Berkers*

**Ruud M.W.J. Berkers**
Modulation of Memory Integration in the Human Brain
Experimental Investigations of Memory Encoding & Consolidation

ISBN 978-94-6375-471-2



Cover design: Myriam Benoit | www.krop.com/benoit_design
Layout and design: Eduard Boxem | www.persoonlijkproefschrift.nl
Printing: Ridderprint BV | www.ridderprint.nl

# Modulation of Memory Integration in the Human Brain

## Experimental Investigations of Memory Encoding and Consolidation

**Proefschrift**

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen

in het openbaar te verdedigen op
woensdag 28 augustus 2019 om 16.30 uur precies

door

**Ruud Martinus Wilhelmus Johannes Berkers**

geboren op 16 juni 1987
te Venray

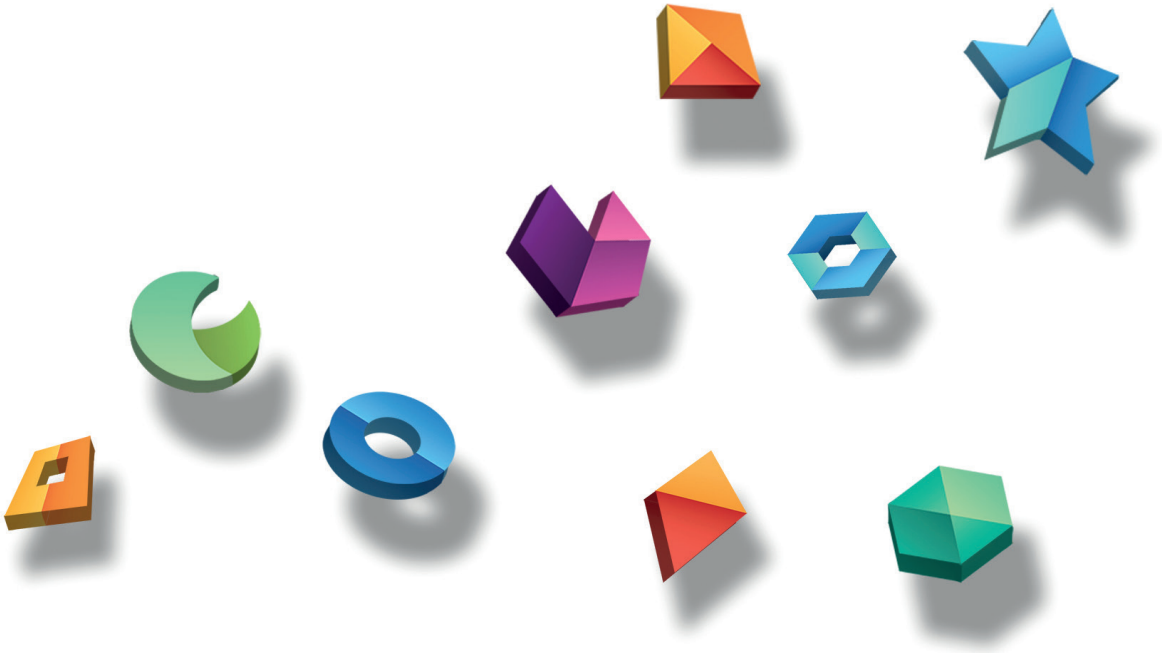
**Promotor**
Prof. dr. G.S.E. Fernández

**Copromotor**
Dr. M.H. van der Linden (Maxima Medisch Centrum)

**Manuscriptcommissie**
Prof. dr. J.M. McQueen
Prof. dr. R.P.C. Kessels
Dr. L.M. Talamini (Universiteit van Amsterdam)

## Table of Contents

*Voor mijn ouders*



# CHAPTER 1

Introduction

## Introduction

At this moment, while I sit in front of my laptop writing this text, nearly a billion seconds[1] have passed by in my lifetime. A billion seconds of experiencing life; seeing, hearing, smelling, tasting, and touching the world around me, to various degrees depending on my level of consciousness (i.e. a significant portion was spent sleeping). During each second of my fully conscious existence, my brain is fed millions of bits of raw data that are then processed to various degrees due to selective deployment of attentional resources (Tononi, 2008). If my brain were to store all information relayed to it, it would contain quadrillions of bits of information. Clearly, this is not what happens. Based on the countless times I embarrassingly fail to remember the names of people that I clearly remember meeting before, or arriving at the grocery store only to have forgotten what I set out to buy there in the first place, I feel safe to conclude that my brain forgets, or never even stored to begin with, a significant amount of all the information it processes.

While I am writing this, I am not merely reproducing the exact information that had been relayed to my brain by my senses at some point in time. Instead, I am creating novel information (this text), expanding beyond information acquired from past perceptions, making novel associations between concepts, logical deductions, analogical relationships and inductive inferences. Thus, our brain is able to go beyond the data available from immediate perception (Tenenbaum, Kemp, Griffiths, & Goodman, 2011), to learn new concepts, acquire language and grasp causal relations. For example, a child can readily learn the meaning of novel words, such as 'dog' and 'candy' from only a few exposures to its exemplars, and is then able to use the word in new situations. These inferences then affect our interpretations of what we are currently experiencing, and they also retroactively affect the manner in which we retrieve our past experiences. Thus, this mnemonic device, our brain, does not blindly and faithfully store all incoming data, rather it selectively filters out irrelevant information, stores incoming information in the context of existing knowledge, and makes inferences that move beyond the immediate perceptual input.

[1] . 963619200 seconds to be precise.

In the end, the cognitive faculty of human memory serves to make us more adaptive, enabling it to make more accurate predictions about its environment and guiding decision-making and behavior that is ultimately beneficial to its survival and reproduction (Nairne & Pandeirada, 2008). To that end, it may not be advantageous to remember all the information that reaches the brain from our senses, but rather to integrate specifically the bits of information that contribute most optimally to making accurate future predictions and avoiding potentially threatening situations (Finsterwald & Alberini, 2014; Shohamy & Adcock, 2010; Smeets, Otgaar, Raymaekers, Peters, & Merckelbach, 2012). Furthermore, it is adaptive for the human brain to make inferences and generalize predictions to situations that extend beyond the limited data available (Anderson, 1991; Schacter, Guerin, & Jacques, 2011; Tenenbaum et al., 2011). The human brain has several such mechanisms to selectively integrate information into the long-term memory store and make inferences.

The goal of this thesis is to answer the central research question: ***How does the human brain selectively integrate information from incoming experience into the long-term memory store?*** This thesis will contribute to the puzzle of human memory, allowing us to approximate a more complete picture of the neural mechanisms of human memory. In this introduction, I will first outline the general theoretical framework delineating how the brain may acquire long-term **knowledge.** Our knowledge store contains a continuously updated model of the world that is built across individual experiences, capturing organizing principles of the relatedness and shared meanings of similar entities. Recent theories and models, as well as experimental findings in humans and animals, have contributed to a preliminary understanding of how such a knowledge store is built. I will specifically discuss the roles of emotional valence, prior knowledge, generalization across episodes, and lastly the selective reprocessing of information during sleep. Throughout the treatment of this material, I will explain how previous research on these mechanisms inspired the research questions and the experimental reports that are presented in **Chapters 2-5.**

## Memory systems in the brain

A discussion of the memory systems of the brain would not be cogent without considering the fate of a patient who came to be known by his initials; 'H.M.'. The seminal case studies performed on this patient substantially impacted conventional thinking about the brain's memory systems. H.M. suffered from intractable epileptic seizures. After attempting several less invasive remedies, neurosurgeon William Beecher Scoville and his team then decided that the only remedy was to surgically resect large parts of the medial temporal lobes, including a removal of the majority of a structure called the **hippocampus** in both hemispheres (Scoville & Milner, 1957). This rather radical procedure served its purpose in partly alleviating the epilepsy. However, a severe and unexpected side effect occurred in the form of a rather profound amnesia. Two subsequent case studies, of patients that underwent temporal lobectomy as well (Penfield & Milner, 1958), added to the overall notion that the hippocampus in particular has an important role in memory.

Noteworthy are those abilities that seemed to be unimpaired by H.M.'s surgery. For instance, language abilities were largely preserved (Kensinger, Ullman, & Corkin, 2001), as well as the ability to learn and retain visuomotor skills (Corkin, 1968; Milner, Corkin, & Teuber, 1968). Later work corroborated this view: several mnemonic abilities seemed to not depend on the integrity of the hippocampus, but instead seemed to depend on extrahippocampal regions. Around the late 1980s into the early 1990s, taxonomies of multiple dissociable **memory systems** (Schacter & Tulving, 1994; Squire & Zola-Morgan, 1988) were proposed, broadly distinguishing between **declarative** memory, e.g. the conscious recollection of facts and events that can be verbally 'declared', and **non-declarative** memories, e.g. memories that are not conscious and cannot be verbally 'declared' (for example, the skill of riding a bicycle). These non-declarative memories are measured as the ability to perform various tasks that tap into motor skills & habits, classical conditioning, perceptual priming and learning, and non-associative learning (i.e. habituation) (Squire, 2004). Declarative memories are the types of memories that are typically impaired in amnesia (Corkin, 2002) and based on studies in humans and animals were deemed dependent on a brain network (see Figure 1.1) including the hippocampus and connected structures in the medial temporal lobe, midline

diencephalon (Aggleton & Brown, 1999), caudate nucleus (Scimeca & Badre, 2012) and other regions in the neocortex, such as the parahippocampal gyrus, medial prefrontal cortex (mPFC) and neocortical association cortices (Eichenbaum, 2000; Preston & Eichenbaum, 2013). The remainder of this thesis will be concerned with primarily the declarative type of human memories.

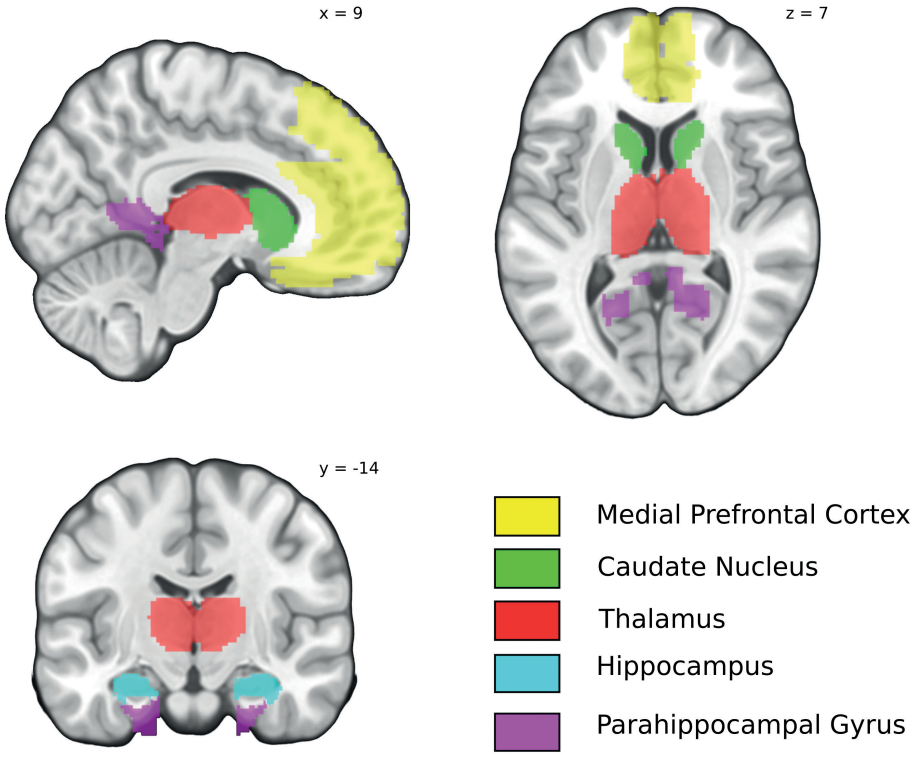



**Figure 1.1. Brain regions involved in declarative memory**

*The upper left panel displays a sagittal slice (x = 9) containing the medial prefrontal cortex, caudate nucleus, thalamus and parahippocampal gyrus. The upper right panel displays an axial slice (z = 7) containing the medial prefrontal cortex, caudate nucleus, thalamus and parahippocampal gyrus. The lower left panel displays a coronal slice (y = -14) displaying the thalamus, hippocampus and parahippocampal gyrus. The regions of interest are based on masks from the IBASPM 71 atlas (thalamus, caudate nucleus & parahippocampal gyrus; Alemán-Gómez et al., 2006), the Anatomy Toolbox (hippocampus; Amunts et al., 2005), and the IBASPM116 atlas (medial prefrontal cortex mask, combining the rectal gyrus, anterior cingulate gyrus, frontal superior medial & frontal mid orbital masks; Alemán-Gómez et al., 2006).*

Further dissociations can be made within the declarative memory system, rooted again in observations of the memory impairments observed in patients such as H.M. Early reports on H.M. characterize his retrograde memory impairments as being temporally graded: memories acquired recently before the surgery were impaired, but remote memories (acquired years or decades prior to surgery) were relatively preserved. For instance, formal testing on H.M.'s memory for events that had occurred prior to his surgery (both public and personal) seemed to show that his retrograde amnesia extended back at least 11 years (Corkin, 1984). An interesting anecdote highlights H.M.'s preserved ability to remember allocentric spatial layouts. After his surgery, H.M. had moved to a new residence where he lived for several years after. At subsequent testing sessions, H.M. was able to draw a correct floor plan as well as provide the address of this home, even though this information was acquired following his lesion (Corkin, 2002). This suggests that H.M. was able to aquire information, albeit slowly over an extended period of time, presumably with the support of extrahippocampal cortical structures. Another patient study also suggested that some forms of declarative memories may not require an intact hippocampus. Here, three patients were described with brain injuries that had occurred at birth or early childhood and resulted in bilateral hippocampal pathology (Vargha-Khadem et al., 1997). Despite a pronounced amnesia for everyday episodes, these patients managed to attend normal schools and attain average levels of speech and language competence, literacy and factual knowledge, suggesting that some forms of declarative memories can indeed be sufficiently supported by extrahippocampal memory regions.

To explain the selective impairments in declarative memory that resulted from hippocampal pathology, two complementary theories were put forward. First, following an initial proposal by Tulving (1972), a distinction was made between **episodic** memory (memory for autobiographical events containing information about the temporal and spatial context of an episode, for example my memory of meeting a friend for dinner one week ago) and **semantic** memory (memory for general facts and personal facts that are independent of spatial or temporal context, for example the name of my mother). In the brain, the hippocampus may be necessary for remembering ongoing episodic life experiences, but not for acquiring factual or general knowledge (Tulving & Markowitsch, 1998).

Second, based on temporal gradients of amnesia found in patients like H.M., a similar temporal gradient in the life of an individual neural memory trace was proposed (Squire, Cohen, & Nadel, 1984). Specifically, upon initial acquisition of an individual memory (**encoding**), it is then processed by neocortical association regions representing individual features that are in turn bound into a coherent memory trace by the hippocampus and related structures in the medial temporal lobes and diencephalon (**storage**), enabling later recovery of the memory trace (**retrieval**). According to this model, a process of prolonged systems consolidation ensues that could last days, months or even decades, during which the contribution of the hippocampus and related structures gradually diminishes while the neocortex is increasingly capable of retrieving the memory trace independently of the hippocampus (Squire, 1992; Squire & Zola-Morgan, 1991). At this point, lesions to the hippocampus would not affect the retrieval of remote memories that had already completed systems consolidation, as they could be sufficiently supported by the neocortex. This **Standard Model of Memory Consolidation** is able to explain the temporally graded pattern of amnesia typically found in H.M.[2]

## Hippocampal-Neocortical interactions and memory consolidation

There are actually two types of memory consolidation in the human brain that need to be considered: initial **cellular consolidation** and protracted **systems consolidation** (Redondo & Morris, 2011). According to the **synaptic tagging and capture hypothesis** (Frey & Morris, 1997), the encoding of a memory trace initially induces the potentiation of synapses, allowing them to become a target for subsequent plasticity-related proteins (PRPs) trafficking that are essential for **long-term potentiation** and **long-term depression** to take place. Long-term potentiation and long-term depression describe how learning takes place in the brain at a cellular level and are responsible for lasting changes in the memory trace. These PRPs may be upregulated depending on the availability of brain-derived neurotrophic factor, and neurotransmitters such as noradrenaline and dopamine. These substances may in turn be upregulated by features of incoming experiences, like novelty (Kemp & Manahan-Vaughan, 2004), reward (Seidenbecher, Reymann,

[2] This pattern of temporally graded amnesia was already reported in the 19th century and formulated in Ribot's law of retrograde amnesia; (Ribot, 1882)

& Balschun, 1997), and physical exercise (van Dongen, Kersten, Wagner, Morris, & Fernández, 2016). Cellular consolidation as described here is a necessary prerequisite for the occurrence of ensuing systems consolidation, and allows memory traces to be reinstated following initial encoding.

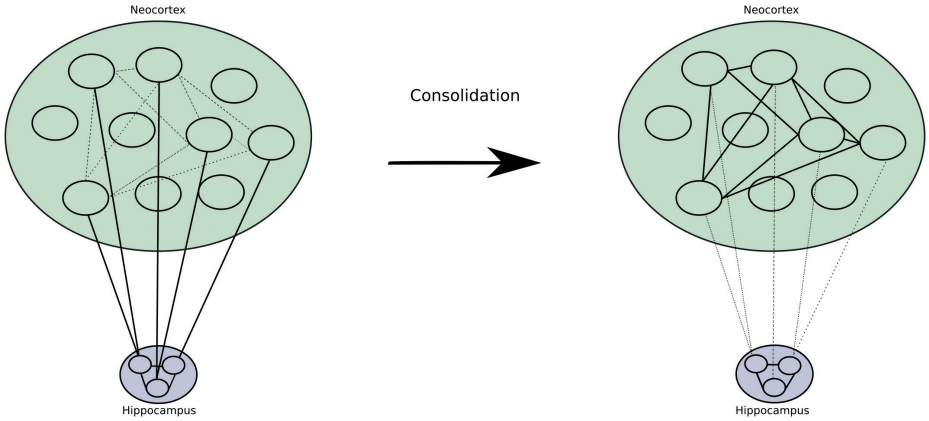



**Figure 1.2. The standard model of memory consolidation in hippocampal-neocortical networks.**

*The standard model of consolidation proposes that information is initially stored simultaneously in neocortical modules and the hippocampus. For a period after learning, systems consolidation takes place, whereby cortical modules are gradually becoming interconnected and the memory trace is stored independently of the hippocampus. This process is slow and takes place across weeks, or months, and may be dependent on sleep (Squire, Genzel, Wixted, & Morris, 2015; Stickgold, 2005).*

The standard model of memory consolidation mainly concerns the second type of consolidation, called systems consolidation. According to this model, memory consolidation involves a reorganization of the neural memory trace, from an initial state of being dependent on the hippocampus, in order to bind distributed neocortically stored features into a coherent trace, to a state where the memory trace becomes gradually more embedded in the neocortex over time (see Figure 1.2). However, the standard model does not clarify the computational principles that may drive this reorganization. These principles may be approximated using **connectionist** models, e.g. artificial neural networks consisting of interconnected networks of nodes, consisting of different basic units such as neurons, synapses and neuronal ensembles (Rumelhart & McClelland, 1986). One prominent

connectionist model proposes **Complementary Learning Systems** situated in respectively the hippocampus and the neocortex (McClelland, McNaughton, & O'Reilly, 1995). After encoding a memory, quick synaptic changes take place in the hippocampal system that then support the subsequent reinstatement of these recent memories in the neocortex. Upon each reinstatement of a particular memory by the hippocampus (as a result of a retrieval cue or prolonged memory replay), neocortical synapses change slightly. Across repeated reinstatements, changes accumulate in the neocortex, leading the memory trace to become more and more embedded into the neocortex. Such an architecture with a fast learning hippocampus and a slow-learning neocortex are deemed necessary to prevent **catastrophic interference,** which is the tendency of a neural network to completely and abruptly forget previously acquired information upon learning novel information. To prevent such catastrophic interference of neocortical memory stores, these models have a gradual information acquisition in the neocortex based on interactions with the hippocampus. Formal models that learn through such gradual and accumulated changes in neocortical connections are demonstrated to be able to discover structure across ensembles of items, as long as the learning of such items is spaced and interspersed with exposures to other items (**interleaved learning**). Thus, while the neocortex slowly learns the structure across experiences, the hippocampus rapidly acquires new memories without disrupting the neocortical structure and subsequently 'trains' the neocortex on these novel memories. The hippocampus is implemented as a sparse, pattern-separating system able to rapidly learn specific episodes, whereas the neocortex is conceived as a distributed, overlapping system that gradually integrates across episodes to extract latent semantic structure (O'Reilly, Bhattacharyya, Howard, & Ketz, 2014). According to this model, the consolidation of a particular memory trace is not necessarily time-dependent, rather it is dependent on the repeated reinstatement of a particular memory and related memory traces (which is often correlated with the passage of time).

The complementary learning systems theory also suggest that those memory traces that are based in the neocortex are those that contain the latent semantic structure accumulated across individual events, whereas the hippocampus codes rather for individual events in their specific temporal and spatial context. This is

congruent with several other theories that have postulated that hippocampal-based memory traces differ in their *format* from neocortical memories (rather than differing in their *age*, as postulated by the standard model of consolidation). For instance, the **Multiple Trace Theory** (Nadel, Samsonovich, Ryan, & Moscovitch, 2000) postulates that episodic memories (memories of events in their temporal and spatial context) are always dependent on the hippocampus for storage and retrieval, whereas semantic or gist-based information (information accumulated across events, lacking their individual temporal and spatial context) can be established independently in the neocortex. Damage to the hippocampus will then only affect episodic information, while semantic information remains intact. It postulates that a specific, context-bound memory trace is *transformed* as it is integrated into the neocortex to become a more schematic, gist-based version of the original memory (Winocur, Moscovitch, & Bontempi, 2010). This model explains the temporally graded retrograde amnesia found in hippocampal patients, as well as unimpaired semantic memories (Nadel & Moscovitch, 1997). Additionally, several other accounts also posit that the storage format of memories is critical in determining whether a memory trace is dependent on the hippocampus (Eichenbaum, Otto, & Cohen, 1994; Hassabis & Maguire, 2009; Mullally & Maguire, 2014; O'Keefe & Nadel, 1978; Palombo, Hayes, Peterson, Keane, & Verfaellie, 2016; Roberts, Schacter, & Addis, 2017; Rudy & Sutherland, 1995; Suddendorf, Addis, & Corballis, 2011; Teyler & DiScenna, 1986). However, these accounts simplify the state of affairs (as models always do) in one crucial respect: they consider memory traces as being encoded on a blank slate. However, memories are always encoded by brains in the context of stored knowledge that has accumulated across previous experiences. I will next consider an account of declarative memory encoding and consolidation which does take prior knowledge into account.

## Schemas and novelty

Already back in 1932, Frederic Bartlett proposed that existing knowledge affects the manner in which new information is processed. He postulated that existing knowledge is stored in memory structures that are constantly developing and modified: so-called **schemas**. He demonstrated this by asking people to verbally recall a story they had read earlier, named *The War of the Ghosts*. By looking at the quality of their reproductions, he concluded that memory retrieval does not take

place by faithfully reproducing information, rather memories are reconstructed at each recollection (Bartlett, 1932). Memory reconstruction is facilitated by each individual's own schema and in some cases the schema biases the reconstruction process as to introduce memory errors. Tulving later recognized that his definition of semantic memory aligned with the traditional definition of schema as a structured network of concepts. According to Tulving, new episodes are experienced, interpreted and encoded through the "prism" of existing semantic networks (Tulving, 1972).

Schemas may thus function as a scaffold that allows related new information to rapidly be integrated. One seminal study by Bransford & Johnson (Bransford & Johnson, 1972) had participants read prose passages that seemed unusual in their semantic meaning. However, whenever they were provided with prior information relevant for understanding the text, their memory for the prose passage significantly improved. Studies have since confirmed that schemas (pre-existing knowledge) may benefit learning (Anderson, 1981; van Buuren et al., 2014; van der Linden, Berkers, Morris, & Fernández, 2017; van Kesteren, Rijpkema, Ruiter, Morris, & Fernández, 2014). Despite the existence of early accounts of the influence of schemas on memory, models of declarative memory consolidation in the brain have for a long time ignored the influence of pre-existing knowledge on the encoding, consolidation and retrieval of a novel memory trace. One common tenet of these models of memory consolidation is that, while initial encoding may occur rapidly, the consolidation of memories into the neocortex is a gradual process that may last up to years (Squire & Alvarez, 1995). However, these accounts fail to take into account studies that suggest neocortical learning may occur rapidly in the presence of schemas. Studies in rats (Tse et al., 2007, 2011) have shown that the presence of an associative schema that is relevant to the learning task can serve as a scaffold, allowing new information to be rapidly incorporated in the neocortex. Here, rats performed a hippocampal-dependent paired associate task, where they learned flavor-place associations in a large arena across many sessions. As such, they formed a putative neocortical schema of the event arena. When new paired-associates were presented, they were learnt quickly. Notably, although the acquisition of new paired associates in the context of an existing schema was hippocampal-dependent, they quickly became hippocampal-independent

as lesioning the hippocampus 48 hr after the acquisition did not impair their retrieval (Tse et al., 2007). In a follow-up experiment, the learning of new paired associates was found to be associated with an up-regulation of immediate early genes (as measured by the availability of two plasticity-related proteins, *Arc* and *Zif268*) in the prelimbic region of the mPFC (Tse et al., 2011). Pharmacological interventions aimed at inactivating this prelimbic area during encoding prevented their subsequent consolidation, indicating that the parallel encoding of memory traces into the prelimbic area is necessary for systems consolidation to occur. This pair of studies thus suggests that on the one hand new associative information can become quickly hippocampally independent during consolidation when congruent with an existing spatial schema, and on the other hand that the encoding of new associative information is also dependent on neocortical regions for later systems consolidation to take place. Human neuroimaging studies have also elucidated the respective roles of the hippocampus and neocortical regions, particularly the mPFC, during schema-dependent learning. Here, increased activity was consistently found in the mPFC during both the encoding (Bonasia et al., 2018; van Kesteren et al., 2013, 2014) and retrieval (Bonasia et al., 2018; Brod, Lindenberger, Werkle-Bergner, & Shing, 2015; van Kesteren, Rijpkema, Ruiter, & Fernández, 2010) of schema-congruent information, while reduced hippocampal activity was found during the encoding (Bonasia et al., 2018; van der Linden et al., 2017; van Kesteren et al., 2013, 2014) and retrieval (Bonasia et al., 2018) of schema-congruent information. This challenges the existence of discrete fast and slow learning systems in the brain, a central tenet of complementary learning systems theory, and suggests that neocortical learning can be fast. Furthermore, schemas can dynamically change the dependence on either hippocampus or neocortex for the acquisition of memory traces.

Theoretically, one could thus differentiate between learning mechanisms for information that is congruent with an existing schema, and information that is incongruent with existing schemas (or novel). The **Schema-Linked Interactions between Medial prefrontal and Medial temporal regions (SLIMM)** framework proposes that when new information is congruent with a schema, a coherent pattern of activity emerges in the cortical network. The resonating and mutually reinforcing activity across the network drives activity within the mPFC, thereby

instantiating the schema. This schema activation facilitates cortical plasticity in the cortical network and the integration of new, related information (van Kesteren, Ruiter, Fernández, & Henson, 2012). In this framework, the mPFC is assumed to inhibit activity within the medial temporal lobe. Other accounts suggest the relationship between these two structures may rather be of a cooperative nature (Gilboa & Marlatte, 2017). The decreased recruitment of the medial temporal lobe (including the hippocampus) for the encoding of schema-congruent information might result in such information being encoded in a manner that is less specific (lacking event-specific temporal and spatial context), and more gist-based (abstracted from its immediate context; Gilboa & Marlatte, 2017). Indeed, it has been found that stronger schema-instatement during encoding is paired with reduced hippocampal activity and retrieval specificity (van der Linden et al., 2017). In turn, when information is highly incongruent with a schema, medial temporal lobe (hippocampal) activity is increased at encoding (van Kesteren et al., 2013, 2014). In such a model, there are thus differing ways for new information to be encoded and consolidated. Completely novel information is encoded initially by the medial temporal lobes in a vivid manner including temporal and spatial context, and gradually, across repeated reinstatements and resulting from hippocampal-neocortical connectivity, integrated into the neocortical memory store. Alternatively, schema-congruent information is assimilated rapidly into the neocortical store, devoid of its episodic details (temporal and spatial context), and integrated in existing knowledge structures.

Findings in humans and animals demonstrate that neocortical networks can quickly acquire new schema-congruent associations. This is inconsistent with the complementary learning systems theory, as originally proposed (McClelland et al., 1995), which holds that the neocortex is slow-learning. In response, the complementary learning systems theory and connectionist models have recently been updated. New simulations show that the artificial network representing the neocortex is indeed able to rapidly acquire new information that is consistent with prior knowledge without inducing catastrophic interference (McClelland, 2013). However, inconsistent new knowledge, when acquired too quickly would induce catastrophic interference. Therefore, inconsistent new information is primarily encoded by the hippocampus and only gradually integrated into the neocortical

memory store. Therefore, the updated model still distinguishes between two complementary learning systems in hippocampus and neocortex, which acquire information in a different manner, but this different learning mechanism is not due to speed of learning (McClelland, 2013). The hippocampus is generally regarded as a **pattern-separating** system of orthogonalized codes capable of rapidly storing specific episodes, in line with empirical findings from animal and human studies (Bakker, Kirwan, Miller, & Stark, 2008; Leutgeb, Leutgeb, Moser, & Moser, 2007; Leutgeb, Leutgeb, Treves, Moser, & Moser, 2004; Yassa & Stark, 2011). If the hippocampus is specialized in pattern separation, the question arises how it is able to support generalization across individual memories, thereby enabling the extraction of a latent semantic structure across episodes that is then stored in the neocortex. A second connectionist model (Kumaran & McClelland, 2012) considers the importance of recurrent interactions between the hippocampus and neocortex that may support efficient generalization across episodes, and the storage of such generalized knowledge in the neocortex. These recurrent interactions between hippocampus and neocortex are implemented in such a way that each subsequent retrieval or **replay** of an individual episode can change the weighting of experience statistics in the generalized neocortical knowledge store. The latest iteration of connectionist models thus implements complementary learning systems in neocortex and hippocampus, with recurrent interactions between hippocampus and neocortex allowing for gradual generalization across episodes, while allowing the rapid integration of new but schema-consistent information (Kumaran, Hassabis, & McClelland, 2016). Recurrent interactions between the hippocampus and neocortex are thus postulated to be driving generalization in the neocortex. The SLIMM-model states that an important role in weighting the consistency of the neocortical schema with new incoming information is played by the mPFC. Therefore, the mPFC acts as a crucial hub through which interactions between distributed neocortical regions representing features of the schema on the one hand, and the hippocampus and its orthogonalized codes for specific episodes on the other hand. The importance of hippocampal-mPFC interactions is supported by the reported increase in mPFC-hippocampal connectivity during the encoding of schema-incongruent information (as found in van Kesteren, Fernández, Norris, & Hermans, 2010; van Kesteren et al., 2014). Additionally, the importance of connections between the mPFC and the hippocampus in this

framework is biologically plausible, as they are anatomically (Öngür & Price, 2000; Saleem, Kondo, & Price, 2008) and functionally (Andrews-Hanna, Reidler, Sepulcre, Poulin, & Buckner, 2010; Ranganath, Heller, Cohen, Brozinsky, & Rissman, 2005) closely interconnected. While this account seems plausible, it is important to consider roles that other regions in the brain may play in this dynamic (see Figure 1.1). For instance, other neocortical regions may be important nodes in the storage of generalized knowledge, such as the angular gyrus and anterior temporal lobes (Binder, Desai, Graves, & Conant, 2009; Rice, Lambon Ralph, & Hoffman, 2015; Wagner et al., 2015). The caudate nucleus of the striatum may also be a fast-learning region, and may in turn gradually train the prefrontal cortex during stimulus-response learning (Pasupathy & Miller, 2005; Seger & Cincotta, 2006) and classification learning (Poldrack et al., 2001). Furthermore, the thalamus, particularly a region called the nucleus reuniens, may be an essential link between the mPFC and the hippocampus (Vertes, Hoover, Szigeti-Buck, & Leranth, 2007), and as such contribute to memory consolidation (Barker & Warburton, 2018; Thielen, Takashima, Rutters, Tendolkar, & Fernández, 2015) and memory generalization (Xu & Südhof, 2013). The potential contribution of each of these regions in the dynamic involvement of parallel learning systems in memory encoding and consolidation will be further considered in the experimental chapters of this thesis.

The models discussed here provide a useful theoretical framework for considering several factors that may influence how new associative information is encoded and consolidated. Particularly, they are helpful in considering how they may tie into the interaction between the mPFC, hippocampus and other brain regions. I will next discuss these factors and brain regions involved in memory encoding and consolidation, and how they inspired our research questions.

## Emotional memory encoding

Memory formation has long been deemed to be influenced in a unique manner by emotion (Hamann, 2001; LaBar & Cabeza, 2006). A well-known phenomenon in this respect are the so-called **flashbulb memories**, extremely vivid memories, which are formed when encountering an event that is deemed highly surprising and perceived to be highly consequential or emotional (Brown & Kulik, 1977).

Examples of these consist of: hearing about a terrorist attack on the news, or more personally: hearing that a relative has passed away. More generally, a significant body of literature has been devoted to an emotional enhancement effect on memory encoding and subsequent retention, finding an enhancement on memory for words, pictures, slide shows or movies (Bradley, Greenwald, Petry, & Lang, 1992; Cahill et al., 1996; Cahill, Babinsky, Markowitsch, & McGaugh, 1995). This emotional enhancement effect has been related to the amygdala (Cahill et al., 1996, 1995; Canli, Zhao, Brewer, Gabrieli, & Cahill, 2000; S. B. Hamann, Ely, Grafton, & Kilts, 1999; Phelps, 2004; Phelps & LeDoux, 2005) and its connectivity to the hippocampus and the wider medial temporal lobe (de Voogd, Klumpers, Fernández, & Hermans, 2017; Dolcos, LaBar, & Cabeza, 2004; Fastenrath et al., 2014; Hermans et al., 2017; Roozendaal & McGaugh, 1997; Smith, Stephan, Rugg, & Dolan, 2006). The enhancement of emotional memories clearly has adaptive value, as emotional events have a high potential future relevance for survival and reproduction.

Studies reporting an emotional enhancement of memory usually test participants on the recognition and recall of items. However, a wider examination of memory functions may reveal that the effect of emotion on memory is variable across study material and testing format. For example, looking more closely at flashbulb memories, it appears that recollections may be highly vivid and confident, but not reliable or consistent across repetitions (Neisser & Harsch, 1992; Talarico & Rubin, 2003). Furthermore, while emotions may improve memory for the gist and perceptual details that are central to the event, memory for peripheral details may be impaired (Burke, Heuer, & Reisberg, 1992). This reduction in memory specificity for emotional events seems linked to an enhancement of attention for the 'center' of an event, at the expense of peripheral detail (Kensinger, Garoff-Eaton, & Schacter, 2007). This effect has been popularized as the **weapon focus effect**, named after the situation where eyewitnesses are less likely to identify a perpetrator when a weapon is present (Steblay, 1992). Here, the weapon as central salient feature is encoded well, but this comes at the expense of identifying the perpetrator's identity (Hope & Wright, 2007), which could be due to the allocation of attention (Christianson, 1992; Easterbrook, 1959; Laney, Campbell, Heuer, & Reisberg, 2004). Several other lab studies have confirmed that while retrieval of the gist or single

item information might be enhanced, retrieval of the source or context of the same emotionally arousing item may be impaired (Kensinger & Schacter, 2006; Mather & Knight, 2008; Pierce & Kensinger, 2011; Rimmele, Davachi, Petrov, Dougal, & Phelps, 2011; Touryan, Marian, & Shimamura, 2007). A direct comparison of memory for items and associations showed that the recognition of negative items is generally improved compared to neutral items, whereas memory for contexts associated with negative items is impaired compared to associative memory for contexts associated with neutral items (Bisby & Burgess, 2014). This effect remains visible when controlling for the allocation of attention, when an item and its associate item are presented in sequence, rather than simultaneously (Madan, Caplan, Lau, & Fujiwara, 2012). Thus, emotional memory might be impaired when testing the associative content, source memory or memory for peripheral detail, thereby rendering the memories less specific to a particular context. This resonates with clinical populations that are characterized with impairments in emotional memory specificity (and overgeneralization of emotional memories), such as appears to be the case in depression and PTSD (Brown et al., 2013; King et al., 2010; Watkins & Teasdale, 2001).

The neural mechanisms underpinning the emotional regulation of memory specificity have not been as widely investigated as the emotional enhancement effect on item memory. However, one consistent feature has been that the mPFC seems to be active in response to emotional valence and arousal (Geday, Gjedde, Boldsen, & Kupers, 2003; Geday, Kupers, & Gjedde, 2007; Phan et al., 2003), and during emotion regulation (Banks, Eddy, Angstadt, Nathan, & Phan, 2007; Quirk & Beer, 2006). Connectivity between the mPFC and amygdala has also been implicated in regulating emotions (Banks et al., 2007; Quirk & Beer, 2006). As discussed before, the connectivity between mPFC and the hippocampus may be implicated in the generalization across episodes, as well as the regulation of the specificity of memory encoding. Indeed, a recent study in mice revealed that mPFC-hippocampal connectivity, mediated by specific thamalic nuclei, regulates the specificity of emotional memories (Xu & Südhof, 2013). However, it remains to be tested how the specificity of emotional memory encoding is regulated in human brains. Reduced specificity of emotional memories might be an adaptive feature of the brain. When encountering a particularly stressful situation (such

as encountering a snake in a particular rocky habitat), it might be helpful to not just avoid the exact context that is predictive of the stressor (the particular rocky formation where you encountered the snake) but also to avoid similar contexts (other rocky formations where a snake might be found).

In the study reported in Chapter 2, we investigated the flipside of emotional memories using a large sample of participants. As such, our sample was highly powered to find subtle effects. More importantly, given the variability of the specificity of emotional memory encoding and impairments in emotional memory specificity in clinical populations, it would be of interest to characterize individual differences in memory performance and relate them to neural differences. These neural differences in activity and connectivity of brain areas can be measured with **functional Magnetic Resonance Imaging (fMRI**, see Box 1). Using these methods, we attempted to answer the question: **How are inter-individual differences in emotional memory explained by activity and connectivity during encoding?**

## Box 1: Functional magnetic resonance imaging

*Functional magnetic resonance imaging (fMRI) is a technique that can be used to indirectly measure brain activity by detecting changes in magnetic properties associated with blood flow and blood oxygenation. Functional MRI is predominantly used for research purposes, but is a specific application of Magnetic Resonance Imaging (MRI), which is a widely used tool in clinical settings. In MRI, a person is placed inside a large magnet, which generates a strong static magnetic field that biases the alignment of hydrogen atoms. These hydrogen atoms act as magnetic dipoles in the body that precess around the external static magnetic field. Additional magnets are then used to generate subtle magnetic gradients in three cardinal directions within the magnetic bore. Radio-frequency pulses selectively excite particular areas inside the magnetic bore (for example the head), thereby temporarily flipping the spin axis of particular hydrogen dipoles out of alignment with the static magnetic field. Once the angle of these precessing dipoles flips back to the aligned state, energy is released inducing a voltage signal that is detected in a receiver (head) coil that oscillates with a certain frequency and a decaying amplitude (free induction decay). Based on this detected signal, the location and intensity of the frequency can be recovered using a Fourier transform. Intensity differences of the MR-signal are found at different tissue*

*types in the brain (such as grey matter, white matter, skull, fat, and cerebrospinal fluid), allowing us to see different tissue boundaries in conventional MR-images. For more information on MRI, see (Brown, Haacke, Cheng, Thompson, & Venkatesan, 2014; Grover et al., 2015).*

*Functional MRI (fMRI) relies on the* ***blood oxygenation-level dependent (BOLD)*** *contrast, which arises due to differing magnetic properties associated with oxygen-rich and oxygen-poor blood (Ogawa, Lee, Kay, & Tank, 1990). Particularly, red blood cells are abundantly present in blood, and contain a protein (hemoglobin) that transports oxygen. In oxygen-rich blood, hemoglobin binds oxygen, causing it to have differing magnetic properties from hemoglobin without oxygen. Using specific acquisition parameters, the BOLD-contrast can be quantified for each volumetric pixel, or voxel, of the MR-image. The BOLD-contrast is closely coupled to neural activation. Neurons extract glucose and oxygen from blood in surrounding capillaries to ensure their energy supply. When a group of neurons becomes activated, more glucose and oxygen are extracted from the local capillaries. A compensatory local response ensues in order to replenish the extracted oxygen, consisting of an increase in blood flow and volume. This response is actually overcompensatory, paradoxically increasing oxygenation of blood in active areas (Bandettini & Ungerleider, 2001; Logothetis, 2003). This* ***hemodynamic response*** *is sluggish, and is reflected in an initial short dip in the BOLD-contrast, followed by prolonged increase peaking at about 6 seconds after initial neural activation, and taking up to 20 seconds or more to return to baseline (****the hemodynamic response function****). The sluggishness of this response ensures that the temporal resolution of fMRI is typically quite low (Boynton, Engel, Glover, & Heeger, 1996). This loss of temporal resolution is compensated by the ability to accurately map of brain activity with a high spatial resolution (see Figure 1.3). For more information see Huettel, Song, & McCarthy (2004).*

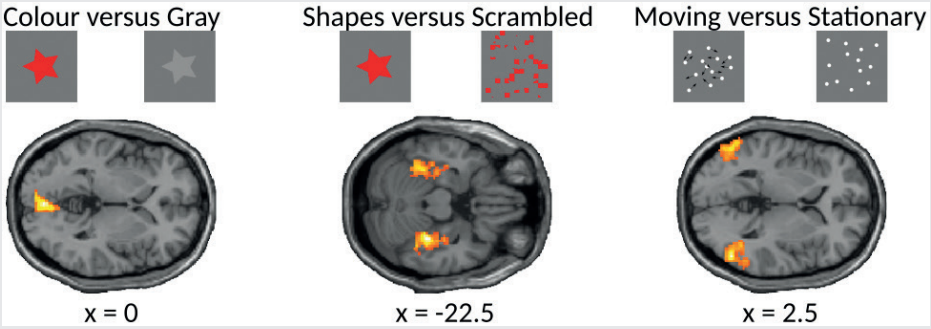



***Figure 1.3. BOLD contrast images of low-level visual features as measured with fMRI.*** *Examples of BOLD contrast images based on the contrasting of functional responses to two different conditions of low-level visual perception (unpublished own data). These contrasts are based on: visual perception of colored versus grayscale objects on a gray background (left panel), visual perception of intact versus scrambled colored shapes on a gray background (middle panel), and visual perception of a field of moving versus stationary white dots on a gray background (right panel). Regions that are color-coded in orange/yellow are regions where an increase in BOLD-signal was detected, after applying a statistical threshold. Regions where higher BOLD-signal is detected are assumed to be more active. The functional contrast images are overlaid on a standard structural image (standard single subject structural image as implemented in SPM, https://www.fil.ion.ucl.ac.uk/spm/).*

## Schemas and memory biases

Schemas can serve as a scaffold to rapidly incorporate new information into the neocortical memory store. According to the SLIMM-model, when information is being encoded that is congruent with an existing schema, it evokes resonating and mutually reinforcing activity across the neocortical network that then drives activity in the mPFC. Such schema instantiation then facilitates cortical plasticity in the cortical network, allowing information to be readily integrated (van Kesteren et al., 2012). In such a situation the brain moves from a hippocampal-based encoding mode, where an event is stored including all its perceptual detail and spatial and temporal context, to a neocortical-based encoding mode, where congruent information is readily incorporated while specific irrelevant details are discarded. This dynamic switching between hippocampal and neocortical encoding modes could be an adaptive feature of memory, allowing the organism to optimize its predictive capacity. In such a mode, highly predictable and congruent events are only partially encoded to the extent that they adapt or add to the existing mental model of the world (the schema). Novel events, those that are

by definition unpredictable, are processed in an event-based hippocampal mode, including all its rich perceptual details and spatial and temporal context. Here, the schema is unable to predict the occurrence event, and thus the schema warrants adjustment. However, as it is not yet known what features are predictive of the event and thus what features should be encoded, the event is stored in all its detail including spatial and temporal context by means of the hippocampus. Gradually, if more similar events are encountered, predictive features can be extracted across these events through recurrent interactions between hippocampus and mPFC (Kumaran & McClelland, 2012) and embedded in a new neocortical memory trace. Future occurrences of this, once novel, event will then become increasingly more predictable and schema-congruent and the organism will become better able to predict its environment.

Despite the general adaptiveness of these alternative memory processing modes, in some circumstances it could lead to the inducement of memory distortions and inaccuracies. The congruency of incoming information with an existing schema could be beneficial for the persistence of the memory trace, but may come at the expense of its specificity (van der Linden et al., 2017). Incorporating schema-congruent information with a reduced involvement of the hippocampus will likely also result in the reduced encoding of irrelevant perceptual detail, thereby potentially rendering the memory trace less specific. Moreover, the existing schema might evoke strong expectations about ongoing events that may in fact proactively interfere with the encoding of the actual event that unfolded. This could in some circumstances result in a false recall of plausible, but inaccurate features of the event. A dramatic demonstration is provided by the **Deese-Roediger-McDermott paradigm**. Here, encoding a list of semantically related words, such as 'rest', 'tired', 'pillow' and 'night', is highly likely to result in the false recall and recognition of the critical lures ('sleep'), which have a strong semantic relation to the other words in the list (Roediger & McDermott, 1995). Encoding a list of words with strong semantic relations to a common concept may create a strong resonating activity across the neocortex, thereby instantiating the 'sleep'-schema. This activity then creates a strong schema-consistent expectation of having encoded the word 'sleep', while actually not having been presented with this word. If the mPFC is important in weighting the influence of the schema, allowing it to exert an influence on

memory encoding, then the mPFC should also be essential in creating such memory errors. Indeed, it has been shown that patients with damage to the ventral mPFC display a reduction in false recall of critical lures on the DRM-task (although false recognition was not found to differ from a healthy control group; Warren, Jones, Duff, & Tranel, 2014). However, findings in patients with lesions are complicated due to the potential for compensatory plasticity and comorbid impairments in other cognitive domains (Murphy & Corbett, 2009). An alternative avenue for research may be provided by **Transcranial Magnetic Stimulation (TMS)**, which is a technique that can be used to temporarily perturb the functioning of a brain region (see Box 2). Particularly, by applying a repetitive magnetic stimulation protocol called **continuous Theta-Burst Stimulation**, ongoing processing in a brain region has been demonstrated to be inhibited for a period of half an hour up to an hour (Huang, Edwards, Rounis, Bhatia, & Rothwell, 2005). In the experiment reported in Chapter 3, we used this technique to down-regulate processing in the mPFC right before participants were required to perform the DRM task. If, as predicted, the mPFC is essential for weighting the influence of the schema on memory formation, and this process results in false memories on the DRM-task, then a down-regulation of the mPFC that persists during the performance of the DRM-task would lead to a reduction in false recall of critical lures. Thus, in **Chapter 3** we attempted to answer the question: **What influence does perturbation of the medial prefrontal cortex have on schema-related memory errors?**

## Box 2: Transcranial magnetic stimulation

*Transcranial Magnetic Stimulation (TMS) is a technique used to temporally interfere with ongoing processing in a particular brain region in a safe and reversible manner (O'Shea & Walsh, 2007). It was first demonstrated in the 80's by Anthony Barker by stimulating peripheral motor nerves that resulted in the activation of effector muscles (Barker, Jalinous, & Freeston, 1985). TMS relies on electromagnetic induction (following Faraday's Law of Induction) to induce a current in the brain using a coil (see Figure 1.4). Specifically, a current flowing through a metal wire (inside the coil) generates a magnetic field surrounding the wire. When the coil is then placed close to a second wire (or other electrical circuit), it induces the flow of an electrical current in the second wire. Similarly, when the coil is placed on the skull, the induced magnetic field passes the skull to generate an electrical current in the brain, thereby*

*stimulating neural activity in the affected brain region, and inducing changes in the immediate neural firing of the area, as well as the overall excitability of the cortex. There are many different coils that can be used, each with unique differences in depth and focality of the induced electrical field (Deng, Lisanby, & Peterchev, 2013, 2014). Furthermore, any given coil can be used to deliver many different stimulation protocols. For instance, the intensity can be varied and is usually calibrated with reference to the motor threshold. The motor threshold is the intensity (measured as % of maximum stimulator output) at which a single pulse delivered to the motor cortex induces an observable response in a corresponding effector muscle. This observable response can either be operationalized as an overt muscle twitch or as a motor-evoked potential measured by electrodes placed on the effector muscle (electromyographic recordings). Furthermore, single pulses can be applied, but also repetitive trains of pulses (repetitive TMS, or rTMS) at varying frequencies. The choice of the exact stimulation parameters depends on the research question. For example, rTMS applied at a frequency of 1Hz results in an inhibition of cortical excitability, whereas a frequency of 5 Hz or higher can produce an increase in cortical excitability (Fitzgerald, Fountain, & Daskalakis, 2006).*

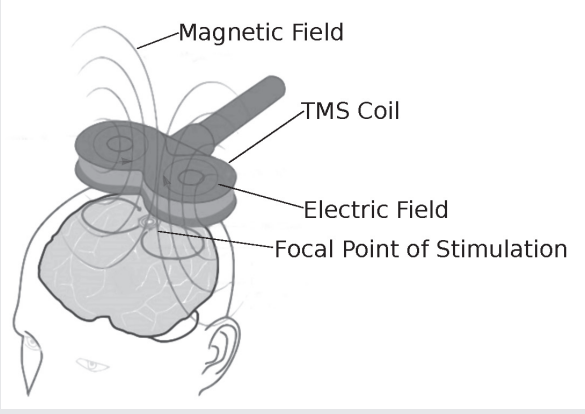



**Figure 1.4. TMS and electromagnetic induction of electrical currents in the brain.**
*Schematic depiction of the generation of electrical currents in the brain by means of a **TMS-coil**. The coil contains a wire through which an electrical current flows that generates a corresponding **magnetic field**. When held close to the head of a study participant it generates an electrical current in the brain (in opposite direction to the current running through the coil). The most common coil is shaped like a figure of 8, as depicted here, and has currents flowing through its two windings in opposite directions. The point where the currents converge in the center of the coil is the **focal point of stimulation** where the induced current in the underlying brain area is the strongest. Adapted from an original figure by Anthony Calvert.*

*The immediate effects of TMS on the cortex are essentially random in nature and add noise to the system, thereby interfering with the organised set of signals that are typically needed to produce cognition, percepts and actions. Stimulation of the motor cortex can produce a muscle twitch, but cannot – as of yet – be used to induce a specific motor action like picking up a cup or kicking a ball. In that sense, online stimulation mostly serves to induce a short-lasting 'virtual lesion', thereby disabling a certain brain region from engaging in ongoing tasks (O'Shea & Walsh, 2007). Stimulation can have 'online' effects simultaneous to or directly following pulse delivery (usually using precisely timed single pulses), but certain rTMS protocols can also have lasting 'offline' effects (Ridding & Rothwell, 2007). Particularly, continuous Theta-Burst Stimulation (cTBS) can produce long-lasting effects of up to an hour, using a relatively short period of stimulation of up to 40 seconds (Huang et al., 2005; Nyffeler et al., 2006). This protocol is therefore ideally suited to be used as an experimental intervention to test the role of a brain region during an extended task block.*

## Acquisition of generalized knowledge

Computational models of learning systems in the brain postulate that the neocortex stores generalized knowledge acquired across repeated episodes. In contrast, the hippocampus contains orthogonalized codes for individual events in their specific temporal and spatial context (McClelland et al., 1995). Across repeated exposures to knowledge elements, recurrent interactions between hippocampus and neocortex may allow for gradual generalization, while pre-existing schemas allow the rapid integration of new schema-consistent information (McClelland, 2013). Furthermore, the caudate nucleus of the striatum may also be a fast-learning brain region, subsequently training slower-learning neocortical brain regions (such as the prefrontal cortex) on associative knowledge (Pasupathy & Miller, 2005; Seger & Cincotta, 2006). Both the striatum and the hippocampus have been demonstrated in patient studies to contribute critically to the initial acquisition of knowledge in the probabilistic classification task (Dalton, Weickert, Hodges, Piguet, & Hornberger, 2013; Holl, Wilkinson, Tabrizi, Painold, & Jahanshahi, 2012; Knowlton, Squire, & Gluck, 1994). Moreover, imaging studies have hinted at complementary roles of the caudate nucleus and the hippocampus during initial learning of (probabilistic) cue-outcome relations in the so-called weather

prediction task (Kumaran, Summerfield, Hassabis, & Maguire, 2009; Poldrack et al., 2001), whereas the generalization of such acquired knowledge was related to initial learning-related activity in primarily the hippocampus and connectivity between hippocampus and mPFC (Kumaran et al., 2009). Generalized knowledge may be stored in different neocortical hub regions, depending on the exact type of knowledge stored, including the mPFC, the angular gyrus, anterior temporal lobes and left inferior frontal gyrus (Barbey, Krueger, & Grafman, 2009; Binder et al., 2009; Rice et al., 2015; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997; Wagner et al., 2015). The hippocampus, caudate nucleus and neocortex may thus have complimentary roles in the acquisition of generalized knowledge.

The cortical acquisition of generalized knowledge may depend on sleep-dependent replay of episodic information to extract regularities and overlap across different events (Lewis & Durrant, 2011; Tamminen, Ralph, & Lewis, 2013), but may also happen within a single learning session with repeated and interleaved learning across trials (Kumaran et al., 2009; Seger & Cincotta, 2006). To tease apart the contributions of hippocampus, caudate nucleus and neocortex across learning, it is necessary to closely track the dynamic unfolding of the updating and accumulation of generalized knowledge about an associative schema across repeated and interleaved learning trials. Therefore, in **Chapter 4**, we had participants perform an iterative learning task inside the scanner where they had to learn associations between words and figures (inspired by Kirby, Cornish, & Smith, 2008). Critically, the associations were organized according to a regular structure that was initially unknown to participants. The participants were then able to extract knowledge of this structure across trials. We fitted learning models to the learning data to be able to track updating and accumulation of knowledge about the associative structure. Model parameters were fitted to trial-by-trial learning activity as measured by fMRI. This allowed us to answer the question: **What brain regions are involved in the accumulation and updating of novel generalized knowledge structures?**

## Sleep & memory reprocessing

Learning models postulate that the neocortex slowly acquires structure across novel experiences, and that the hippocampus, in contrast, rapidly acquires traces

of new experiences without disrupting this structure and subsequently trains the neocortex on newly acquired regularities. This training of the neocortex through hippocampal-neocortical interactions may occur as a result of repeated exposures to particular learning experiences during awake experience, allowing overlap and generalized features to be extracted across experiences (see previous paragraph). This process may be particularly mediated by the mPFC and its interactions with the hippocampus (Kumaran et al., 2009; Zeithamova, Dominick, & Preston, 2012). Additionally, rodent studies have shown that awake learning experiences may subsequently be replayed during **slow-wave sleep** in the hippocampus (Carr, Jadhav, & Frank, 2011; Carr, Karlsson, & Frank, 2012; Skaggs & McNaughton, 1996) in concert with neocortical areas (Euston, Tatsuno, & McNaughton, 2007; Ji & Wilson, 2007; Siapas & Wilson, 1998) and the striatum (Lansink, Goltstein, Lankelma, McNaughton, & Pennartz, 2009). When replaying experiences, there might be an overlap across replayed experiences. Precisely these shared aspects of experiences might be strengthened, thereby training the neocortex on these statistical regularities. As such, sleep-related memory replay could be essential for schema formation in the neocortex, as well as the assimilation of novel information into these schemas (Lewis & Durrant, 2011; Stickgold & Walker, 2013). This is congruent with reports that slow-wave sleep is not only crucial for memory consolidation in general (Fowler, Sullivan, & Ekstrand, 1973), but also the integration of new information into neocortical schemas (Tamminen et al., 2013) and the learning of statistical regularities (Durrant, Taylor, Cairney, & Lewis, 2011).

Slow-wave sleep, also referred to as deep sleep, is classified as stage 3 and 4 of non-rapid eye movement sleep. It is operationalized as any epoch of 30s of sleep that contains 20% or more delta-wave activity (Iber, Ancoli-Israel, Chesson, & Quan, 2007; Schulz, 2008). During this sleep stage, slow oscillations may control global brain-wide polarization, facilitating the occurrence of **thalamic spindles** during 'up-states' characterized by cortical depolarizations (Mölle, Marshall, Gais, & Born, 2002). These spindles are oscillatory signals that are generated by the thalamus, and have in turn, been related to overnight memory retention (Schabus et al., 2004). Spindles have in rodents been shown to be tightly coupled to the occurrence of **hippocampal sharp-wave ripples** (Siapas & Wilson, 1998),

which are themselves related to memory consolidation (Girardeau, Benchenane, Wiener, Buzsáki, & Zugaro, 2009). Furthermore, intracranial recordings of local field potentials in humans (Staresina et al., 2015), and optogenetic studies in mice (Latchoumane, Ngo, Born, & Shin, 2017) have demonstrated a tight coupling between these three cardinal oscillations, which may provide a substrate for precisely timed information transfer between hippocampus and neocortex (Sirota, Csicsvari, Buhl, & Buzsáki, 2003). Furthermore, recordings of neurons in the mPFC in rats have demonstrated the occurrence of learning-dependent replay during slow-wave sleep of those patterns that were initially observed during rule learning (Peyrache, Khamassi, Benchenane, Wiener, & Battaglia, 2009). This neuronal replay occurred simultaneously to hippocampal sharp-wave ripples, which is consistent with the notion that these oscillations are electrophysiological correlates of memory replay in hippocampal-thalamic-neocortical circuits.

One way to investigate memory replay during sleep is to externally elicit the reactivation of memory traces (during sleep) by presenting mnemonic cues, (so-called **targeted memory reactivation)**. For instance, smells can be paired with particular learning experiences and subsequently presented as cues to participants during slow-wave sleep. As such, it has been shown that targeted memory reactivation with odors during slow-wave sleep can promote memory consolidation of material acquired previously in the context of the smell. Furthermore, the cue smell presentation itself was found to be related to the increased activation of the hippocampus and posterior neocortical areas as measured with concurrent fMRI (Diekelmann, Büchel, Born, & Rasch, 2011; Rasch, Büchel, Gais, & Born, 2007). However, smells are not ideally suited for the selective cueing of individual memories. Auditory cues can also be used to reactivate individual memories during slow-wave sleep, which has been related to memory consolidation of those specific memories that were previously cued during slow-wave sleep (Cairney, Durrant, Hulleman, & Lewis, 2014; Rudoy, Voss, Westerberg, & Paller, 2009). If cued memory reactivation during slow-wave sleep is akin to endogenous memory replay during sleep, it should take place in concert with a coordinated interplay between the hippocampus, thalamus and neocortical brain regions. As such, auditory cues that were paired with learning material can be presented again to participants in the scanner during slow-wave sleep.

This allows a comparison of patterns of connectivity and activity in the brain during the presentation of auditory cues with the presentation of sounds that were not previously paired with learning materials. The presentation of auditory cues to participants in deep sleep inside a loud scanner environment is difficult. Notwithstanding such practical difficulties, one study found increased activation of parahippocampal cortex during the presentation of such cue sounds, and increased connnectivity of this parahippocampal region with posterior visual cortex regions that represent memory features (van Dongen, Takashima, et al., 2012). However, memory reactivation during slow-wave sleep might be paired with a complex reconfiguration of synchronous brain activity and connectivity across multiple brain regions. Therefore, it might be helpful to capture changes in connectivity patterns during cued memory reactivation in slow-wave sleep by looking across the entire brain. Graph theoretic measures can be used for the purpose of characterizing local and global topological features of the brain (see Box 3). Therefore, in **Chapter 5** we used graph theoretic measures to re-analyze the data reported in van Dongen et al. (2012) to capture shifts in whole-brain connectivity in response to cued memory reactivation during deep sleep. As such, we hope to answer the research question: **What brain connectivity changes are induced by cued memory reactivation during sleep and are they related to memory stabilization?**

## Box 3: Using graph theory to characterize brain networks

*The brain is a large, complex, and widely interconnected structure. Conventional approaches of measuring functional connectivity often only consider two or a limited set of brain regions (Biswal, Zerrin Yetkin, Haughton, & Hyde, 1995; Friston et al., 1997; Friston, Harrison, & Penny, 2003), thereby inevitably ignoring the complexity of connectivity patterns across the entire brain network. Complex network analysis is an approach that can be used to quantify local and global properties of connectivity within such complex systems (Bullmore & Sporns, 2009; Rubinov & Sporns, 2010). It is based on the mathematical study of graphs, which can be traced all the way back to 1736, when Euler formulated the famous 'problem of the seven bridges of Königsberg'. A graph is a mathematical structure that is a model of pairwise relations between objects (Newman, 2018). A graph is made up of nodes, which are connected by edges. The connecting edges may be directed, but can also be undirected. The brain can*

*be conceptualized as a graph, by including anatomical brain regions as nodes that are connected by anatomical connections or functional connectivity as measured by correlations of functional timecourses (Bassett et al., 2009; Bassett & Bullmore, 2006; Stam & Reijneveld, 2007). The brain regions, or nodes, can be defined based on gross anatomical features, such as the AAL atlas (Tzourio-Mazoyer et al., 2002) or finer anatomical units such as voxels. Complex network analysis can be used to quantify complex brain networks with easily computable measures and can be used to explore differences in network topology between task states (Backus, Bosch, Ekman, Grabovetsky, & Doeller, 2016; Ekman, Derrfuss, Tittgemeyer, & Fiebach, 2012), or between healthy and patient populations (Boersma et al., 2013; van den Heuvel et al., 2013).*

*Functional brain networks can be constructed based on resting-state fMRI data using raw functional timecourses (Deco, Jirsa, & McIntosh, 2011; Dosenbach et al., 2007), or task fMRI data (Cao et al., 2014; Ekman et al., 2012) based on activity extracted across single-trial activity estimates (beta-timeseries analysis; Rissman, Gazzaley, & D'Esposito, 2004). These activity timecourses are extracted for every brain region, or node, and the timecourses for every node are correlated with every other node to construct a functional connectivity matrix. This matrix can be binarized, where every connection above a certain threshold is included as a 1, and every other connection is set to 0. In addition, weighted matrices can also be used. Here, a stronger functional connectivity between two brain regions corresponds to a shorter connecting edge. Several measures can be taken to quantify, for example, network integration (or conversely, network segregation) of the brain as a whole, network centrality of individual brain pathways or brain regions, resilience of the network to insult or lesions, and local anatomical circuitry. Here, I will further describe measures capturing network integration of the brain as a whole, and network centrality of individual brain regions*

*The most basic network measure is the nodal degree, quantified as a node's total number of connections in case of a binary network, or the sum of all weighted connections in case of a weighted network (Rubinov & Sporns, 2010). The degree of a node reflects the importance of a node for the network, or how central it is to the wider network ('degree centrality'). However, merely looking at a node's degree*

*may not be the best indicator of whether a brain region constitutes a hub in the wider brain network (Power, Schlaggar, Lessov-Schlaggar, & Petersen, 2013). Hubs integrate information across the entire brain as a result of the strength, number, and positioning of their connections within a network. A region with a high degree centrality may in fact only be densely connected within a single brain system, while not being connected to other brain systems at all. To find regions that have connections across brain systems it is important to determine the community structure. Such community structure can be defined by dividing a network into separate modules: groups of densely interconnected regions of differing sizes and composition (see Figure 1.5). The network is divided into groups of nodes in such a way as to maximize the number of within-module connections and minimize the number of between-group connections. The extent to which a network can be subdivided into such modules is quantified with the modularity statistic. The optimal community structure is usually estimated using optimization techniques (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008; Newman, 2006). A brain characterized by a high modularity statistic can be subdivided into many modules, where specialized processing is likely to occur within densely interconnected groups of brain regions. Such a brain displays a high level of functional segregation. Conversely, a brain with a low modularity statistic can be subdivided into few modules and is thus highly integrated.*

*Having obtained the optimal community structure, each region's involvement in inter-modular connections can be determined. The participation coefficient measures a node's within-module connections versus its between-module connections (see Figure 1.5). Nodes with a high within-module degree, but low between-module degree, likely facilitate modular segregation (provincial hubs). On the other hand, nodes with a high participation coefficient are characterized by a high number of between-module connections relative to its within-module connections, and these nodes are therefore likely to facilitate global intermodular integration (also called connector hubs). Regions with a high participation coefficient thus likely constitute critical hubs in the wider brain network.*

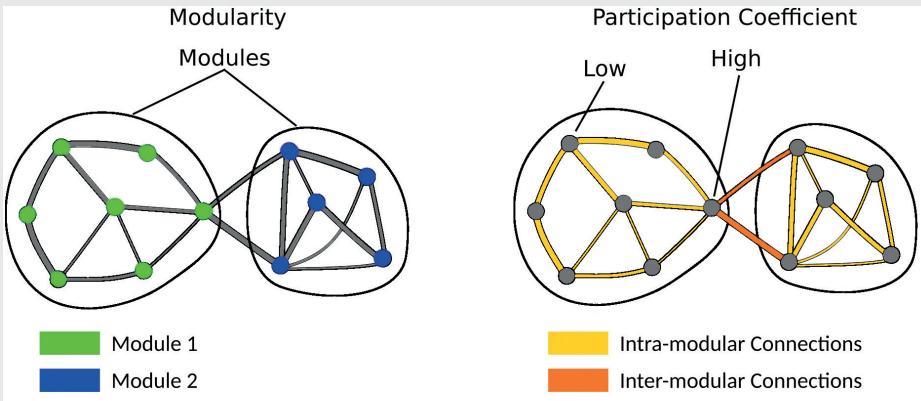
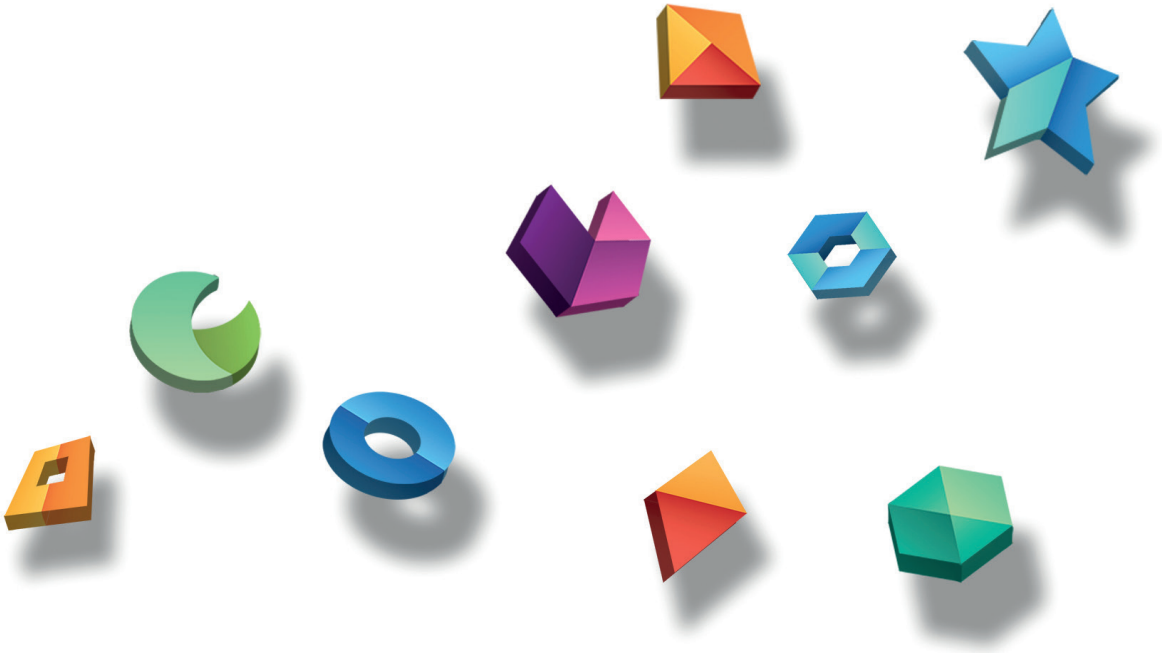



**Figure 1.5. Modularity and participation coefficient as network measures.**
*Schematic depiction of a simple network of nodes connected by vertices. This schematic network may be subdivided into two submodules (see left panel) consisting of a high number of within-module connections and two between-module connections (see right panel). Nodes with a high participation coefficient have a relatively high number of between-module connections versus its within-module connections. Figure adapted from Rubinov & Sporns, 2010.*

## Thesis outline

The next four chapters (**Chapters 2-5**) will continue with a description of the four experimental investigations that each address the research questions introduced in the previous sections. After having described these experiments, I will discuss in **Chapter 6** how each of these respective studies have answered their specific research questions in the General Discussion. Furthermore, I will discuss how my studies have contributed to a more complete answer to the central research question, which as a reminder, was formulated as: ***How does the human brain selectively integrate information from incoming experience into the long-term memory store?*** In doing so, I will revisit the theoretical framework previously outlined and discuss how it may be modified based on the experimental studies presented.



# CHAPTER 2

## Medial prefrontal-hippocampal connectivity during emotional memory encoding predicts individual differences in the loss of associative memory specificity

This chapter is based on:

Berkers, R.M.W.J., Klumpers, F., & Fernandez, G. (2016). Medial prefrontal-hippocampal connectivity during emotional memory encoding predicts individual differences in the loss of associative memory specificity. *Neurobiology of Learning and Memory, 134, 44–54.*

## Abstract

Emotionally charged items are often remembered better, whereas a paradoxical loss of specificity is found for associative emotional information (specific memory). The balance between specific and generalized emotional memories appears to show large individual differences, potentially related to differences in (the risk for) affective disorders that are characterized by 'overgeneralized' emotional memories. Here, we investigate the neural underpinnings of individual differences in emotional associative memory. A large group of healthy male participants were scanned while encoding associations of face-photographs and written occupational identities that were of either neutral ('driver') or negative ('murderer') valence. Subsequently, memory was tested by prompting participants to retrieve the occupational identities corresponding to each face. Whereas in both valence categories a similar number of faces were labeled correctly with 'neutral' and 'negative' identities, (gist memory), specific associations were found to be less accurately remembered when the identity was negative compared to neutral (specific memory). This pattern of results suggests reduced memory specificity for associations containing a negatively valenced component. The encoding of these negative associations was paired with a selective increase in medial prefrontal cortex activity and medial prefrontal-hippocampal connectivity. Individual differences in valence-specific neural connectivity were predictive of valence-specific reduction of memory specificity. The relationship between loss of emotional memory specificity and medial prefrontal-hippocampal connectivity is in line with the hypothesized role of a medial prefrontal-hippocampal circuit in regulating memory specificity, and warrants further investigations in individuals displaying 'overgeneralized' emotional memories.

## Introduction

Emotional events can be pervasively engrained in memory, as demonstrated by vivid recollections of flashbulb-memories (Brown & Kulik, 1977) or intrusive memories of patients with post-traumatic stress disorder (Brewin, Gregory, Lipton, & Burgess, 2010). Indeed, an emotional enhancement effect is reliably found experimentally when probing memory for items like faces, objects, scenes, words or movie clips that are charged with a negative emotional valence compared to neutral items (Bradley et al., 1992; Cahill et al., 1996, 1995; Canli et al., 2000; Hamann et al., 1999; Talmi & Moscovitch, 2004). However, a somewhat different picture emerges when memory is tested beyond isolated items, probing memory for associated items or spatiotemporal context. While memory for negative items themselves is enhanced compared to neutral items, memory for associated items or associated context is impaired (Bisby & Burgess, 2014). Even when an impaired memory for associative detail is found, the subjective sense of vivid recollection can be increased (Rimmele et al., 2011).

The paradoxical modulatory effect of emotion on memory is reminiscent of the so-called 'weapon-focus' effect reported in the eye-witness literature, whereby a perceived item of negative valence (such as a weapon) impairs witnesses' ability to identify the perpetrator carrying the gun (peripheral associated information) (Christianson & Loftus, 1991; Migueles & Garcia-Bajos, 1999; Steblay, 1992). It has thus been argued that emotional valence might enhance the likelihood that the theme or gist of an event is remembered at the expense of memory for specific details (Adolphs, Tranel, & Buchanan, 2005). Recently, it has been postulated that emotionally arousing items attract particular attention, thereby enhancing binding of its constituting elements. At the same time, the association of the central object with contextual information and other objects is weakened (Mather, 2007). These two tenets are not contradictory, but are rather shown to complement each other in explaining the emotional memory paradox. When comparing encoding of negative versus neutral material, detailed memory is preserved for particularly those objects that are central, whereas memory for the non-emotional background becomes less detailed (Kensinger et al., 2007). Similarly, when encoding emotional material, the increased subjective sense of recollection is found to be related to an enhanced memory for the what, where and when of a specific emotional

item, whereas peripheral details and associations to other items occurring at the same time are poorly remembered (Bisby & Burgess, 2014; Rimmele et al., 2011; Rimmele, Davachi, & Phelps, 2012). While an emotionally arousing picture impairs memory for the background pattern, it does not impair item recall, item recognition, or location memory of another central picture or object and its features (Erk et al., 2003; Mather, Gorlick, & Nesmith, 2009; Touryan et al., 2007). Emotional items also impair memory for the specific association amongst them. For instance, one study found that recognition memory for negative word-pairs was found to be impaired compared to neutral or positive word pairs, and this was expressed primarily as an increased false alarm rate for re-arranged negative word pairs (Pierce & Kensinger, 2011). Notably, the hit rate for negative words demonstrated decreased forgetting across the one week consolidation period, but this finding was paralleled by an increase in the false alarm rate. Thus, while the emotional valence boosts item recall and item recognition, it reduces memory specificity for associations and background context.

When investigating the brain basis of the emotional modulation of memory, a distinction is thus warranted between memory for the central item and memory for associated information and spatiotemporal context. Corroborating this distinction is the finding that damage to the amygdala, a brain region deemed to be essential for the emotional modulation of memory (McGaugh, 2004; Phelps, 2004), impairs gist memory while retrieval of associated detail is preserved (Adolphs et al., 2005). Individual differences in emotion-modulated item memory can be related at the neural level to activity in the amygdala and other medial temporal lobe (MTL) regions, as well as their mutual connectivity (Dolcos et al., 2004; Hamann et al., 1999; Kilpatrick & Cahill, 2003; Murty, Ritchey, Adcock, & LaBar, 2010; Ritchey, Dolcos, & Cabeza, 2008). The neural mechanisms underlying the emotional regulation of memory specificity are, however, less clear. In general, there is a wide set of regions implicated in the retrieval of associative detail in both cued recall and source memory tasks, including the medial prefrontal cortex (mPFC), posterior midline, bilateral parietal and medial temporal regions (Hayama, Vilberg, & Rugg, 2012; Vilberg & Rugg, 2014). Particularly, the mPFC has been found to respond also to emotional valence and arousal (Geday et al., 2003, 2007; Phan et al., 2003), to social and self-referential processing (Gusnard,

Akbudak, Shulman, & Raichle, 2001; Mitchell, Banaji, & Macrae, 2005; Mitchell, Neil Macrae, & Banaji, 2005) and emotion regulation (Banks et al., 2007; Quirk & Beer, 2006). Connectivity between the mPFC and amygdala have been found to be important for regulating emotions (Banks et al., 2007; Quirk & Beer, 2006). Recent findings have converged in implicating the mPFC also in the integration of information in preexisting knowledge structures, so-called schemas that appear to be memories that are generalized over several episodes (Lewis & Durrant, 2011; van Kesteren et al., 2012). Thus, it can be suggested that the mPFC plays a critical role in assimilating generalized memories that are less specific to an individual episode. A recent study in mice has added critical support for this hypothesis by revealing that mPFC-hippocampal connectivity mediates the specificity of emotional associative memories (Xu & Südhof, 2013). However, it remains to be tested whether this interaction is also involved in regulating emotional associative memories in humans.

Here, we investigate the neural mechanisms related to the modulation of associative memory specificity due to emotional valence. Individual differences in neural mechanisms might be particularly interesting when informing future research into extreme impairments in emotional memory specificity found in clinical populations (Foa, Gilboa-Schechtman, Amir, & Freshman, 2000; Moradi, Taghavi, Neshat-Doost, Yule, & Dalgleish, 2000; PWatkins, Vache, Verney, & Mathews, 1996). We explore these individual differences here in a healthy population using a memory encoding task where faces needed to be associated with occupational identities. This task has been frequently used in the literature to elucidate individual variation in associative memory (Dominique & Papassotiropoulos, 2006; Erk et al., 2010, 2011; Werner et al., 2009). Here, we added a manipulation of emotional valence of the identity labels to allow us to look at individual differences in emotional associative memory. Specifically, a large group of young healthy men were scanned while encoding associations of individual faces and either neutral (e.g., 'driver') or negative (e.g., 'murderer') occupations, and were asked later to identify the particular neutral and negative occupational identities associated with each particular face. Broadly, this laboratory task resembles a line-up situation where a victim needs to identify a villain from a range of other identities. Memory performance is operationalized on a general, gist-level (a face is associated with

an identity from the correct valence-category but not the exact identity) and a specific, detailed level (the specific face-identity association is remembered correctly). A loss of memory specificity can be expected for those associations that consist of negative occupational identities, potentially mediated by interactions of the hippocampus with the mPFC.

## Materials and methods

### Participants

One-hundred-twenty young healthy male volunteers in the age range of 18 – 31 (mean age 21.9; SD = 2.63) provided informed consent to participate in the study. All subjects were right-handed and pre-screened to exclude a history or current status of psychiatric, neurological or endocrine disorder, and to exclude the consumption of illicit drugs or medications affecting the central nervous or endocrine systems at any point over the past six months. The study was conducted in accordance with guidelines of the local ethics committee (Commissie Mensgebonden Onderzoek region Arnhem-Nijmegen, The Netherlands) and the declaration of Helsinki. To ensure that all subjects in the final sample understood and performed the task correctly, we removed those participants from further analysis that performed extremely poorly. The threshold was set at two items correct, where the items were counted correct when both valence+identity was correct. Two items corresponds to the number of items that can be expected to be guessed correctly if subjects were fully aware of the valence of each face and completed every item of the test (32 faces, 16 negative occupations, 16 neutral faces, 1/16 chance of having the occupation correct per item, 32 * 1/16 = two items). Using this threshold resulted in a final sample of 102 participants for further analysis.

### Procedure

The experiment was embedded in a larger study that entailed two visits of each participant, separated by on average two weeks (minimum of five days). Here, we focus on data from one task, while controlling for order-effects.

**Experimental paradigm**

Participants performed the emotional face-occupation association task inside the MRI scanner, where they were presented simultaneously with pictures of faces and written words describing occupations that were either of neutral (e.g. driver) or of negative valence (e.g. murderer). To encourage the subjects to engage in deep encoding, the instruction during the associative learning of the face-occupation pairs was to imagine the person in the picture matching the description underneath the photo. Participants then indicated with a button press whether they could imagine the person being labeled with the associated description. Furthermore, they were instructed to remember these associations for a subsequent memory test.

Thirty-two pairs were used per subject. Half of these pairs included an identity word that was of neutral valence, whereas the other half included a negative identity word. The unique combinations of these 32 faces and 32 occupational identities were counterbalanced across subjects. The words in the negative and neutral categories were matched in terms of word length and frequency, but differed in terms of rated valence and arousal as shown by independent behavioral pilots. Furthermore, the two sets of words did not differ significantly in terms of their semantic similarity (see Table 2.1).

**Table 2.1.** Characteristics of neutral and negative words.

| Word characteristic | Mean neutral | Mean negative | P-value |
|---|---|---|---|
| Length | 8.88 | 8.34 | 0.26 |
| Frequency | 111.81 | 91.69 | 0.53 |
| Valence | 5.38 | 2.82 | <0.001* |
| Arousal | 2.26 | 4.41 | <0.001* |
| Semantic similarity | 0.30 | 0.29 | 0.08 |

*Word characteristics were compared for neutral and negative word lists. The word length was determined by the number of letters contained in the word. The frequencies of these words in the Dutch language were obtained from the Celex-database, and its units consist of the frequency per million words in the whole corpus (Baayen, Piepenbrock, & Rijn, 1993). The ratings of valence and arousal of each word were obtained by an independent behavioral pilot in a separate group of subjects (N = 18). The valence of each word was rated on a scale from 1 to 9 where a lower score equals higher emotional valence, whereas the evoked arousal of each word was rated on a scale from 1 to 9 were a higher score equals higher arousal. Corpus-based pairwise word similarity measures were obtained from Cornetto, a lexical-semantic database for the Dutch language (Vossen et al., 2013). Semantic similarity was calculated using Lin's information-theoretic measure of similarity (Lin, 1998), resulting in word-unique values ranging from zero (completely unrelated) to one (identical). Statistical comparisons were made using a two-sample t-test, except for semantic similarity, where Welch's two-sample t-test was used.*

The pairs were presented in study blocks of 24 seconds, during which four pairs were presented for six seconds each. There were four blocks of face-occupation associations with a neutral valence, and four blocks that included negative identity words. The experiment included also three blocks of a perceptual control task, where participants were presented with three blocks of four trials, each requesting a perceptual judgment. Here, subjects were presented with head contours (grey silhouettes against a colorful background), and had to indicate with a button press whether the left or right ear was lower in the picture. The sequence of blocks was counterbalanced across subjects, and blocks of the three conditions were presented in alternating order such that each block was always followed by a block from a different condition. Each block was separated by a 2 second interval, during which participants were cued with the task of the ensuing block ('associate face' was presented for encoding blocks and distance ear' for perceptual blocks). All blocks were presented in a single scanner run.

After a delay period, participants were tested outside the scanner on their memory for the associations. While there was variance between subjects in the delay between encoding and recall due to practical reasons (mean: 47:31 mins, SD = 09:44 mins, range: 11:40 – 1:00:53), the delay period did not significantly predict individual subject performance on behavioral memory measures reported here. Furthermore, given that we recorded the delay for each subject, we insured it would not bias the reported brain-behavior correlation by taking it into account as a control variable in a partial correlation analysis. At test, subjects were given lists of both the face-pictures and written occupational identities that were presented at encoding, and were asked to recall the associations and write the identities underneath the faces (cued recall task). They were not given any novel faces or occupational identities as lures. Furthermore, subjects were instructed to avoid guessing and to respond with those associations that they recalled with confidence (see Figure 2.1).

**Behavioral analysis**

Behavioral measures were analyzed in IBM SPSS Statistics 21. Behavioral performance on the cued recall task was scored by categorizing items into four bins based on response accuracy. Besides the 'misses' (items where no response was given), the responses could be divided into those that have both an incorrect identity and an incorrect valence ('incorrects'), those that have an incorrect identity yet belong to the correct valence ('valence-only correct') and those that have a correct identity ('valence+identity correct'). A measure of memory specificity was then calculated for neutral and negative valence categories separately by dividing the 'valence+identity correct' items by the total 'valence-correct' items (the latter also including the valence-only correct items). The difference between memory specificity scores for each valence was calculated to assess whether the negative valence would lead to the predicted loss of specificity. Subsequently, Student's *t*-tests were performed to determine differences in performance between the negative and neutral valence conditions (paired-sample *t* test). For all tests $\alpha$ was set at 0.05.

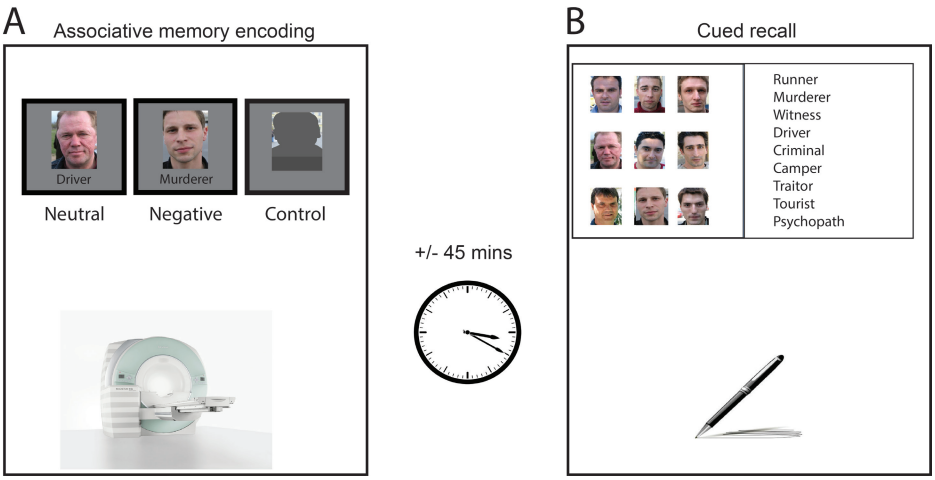



**Figure 2.1. Task setup.**

*The task consisted of two phases, namely an associative memory encoding phase, and a cued recall phase.* ***A,*** *In the associative memory encoding phase, participants were presented in the scanner with pictures of faces paired with a written identity (in Dutch, but English translations shown here), divided into blocks of identities with a neutral valence, and identities with a negative valence. Participants indicated with a button press whether they could imagine the written identity belonging to the face in the picture. In the control task blocks, participants were presented with a grey head silhouette, and responded with a button press whether the left or right ear was lower in the picture.* ***B,*** *In the cued recall phase, participants were presented with a list of all face pictures and a list containing all written identities (subsets are shown here). Participants had to recollect the original face-identity pairs by writing the identity underneath the corresponding picture.*

### MR data acquisition

Participants were scanned using a 1.5 Tesla Magnetom Avanto MR-scanner (Siemens, Erlangen, Germany) equipped with a 32-channel phased array head coil (MRI Devices). For blood oxygen level-dependent (BOLD) functional magnetic resonance imaging (fMRI) images, we used a T2*-weighted gradient echo planar imaging sequence to collect a series of 128 images during the task run with the following parameters: repetition time (TR): 2.34 s, echo time (TE): 35 ms, 32 oblique transverse slices, flip angle: 90°, slice matrix size = 64 × 64; slice thickness = 3.5 mm; slice gap = 0.35 mm; and field of view (FOV) = 212 × 212 $mm^2$, and voxel size: 3.3 × 3.3 × 3.5 mm. To ensure reaching a steady-state condition, the first five scans were discarded. Additionally, 3D magnetization-prepared rapid gradient echo (MPRAGE) anatomical T1-weigthed images were acquired

for normalization purposes (176 slices, 1.0 mm isotropic resolution, TR = 2730 ms, TE = 2.95 ms).

**MR data analysis**

Functional MRI data were analyzed using SPM8 (UCL, London: http://www.fil.ion.ucl.ac.uk/spm), following standard preprocessing procedures. Specifically, images were corrected for slice-timing and realigned to account for three-dimensional motion. After co-registration of the structural and the functional images, all functional images were normalized into standard stereotactic space using the T1 (Montreal Neurological Institute [MNI] 152-template). Smoothing was performed with a Gaussian kernel of 8 mm full-width at half-maximum.

Next, single subject general linear models were constructed with 24-s boxcar regressors modeling task blocks separately for neutral valence, negative valence and control task. Regressors were temporally convolved with the canonical hemodynamic response function. The six movement parameters were included as covariates to model potential movement artifacts. Contrast parameter images were generated at the single subject level for negative and neutral encoding versus the perceptual control task. These individual parameter estimate maps were statistically scrutinized at the second level. In a factorial ANOVA emotion type was added as within-subject factor and session order (whether the session was the first or the second session) as between-subject factor.

Effective connectivity between regions was determined using psycho-physiological interactions (PPIs) as implemented in SPM8. PPI-analyses assess differences in co-activation of a seed region (physiological factor) with the rest of the brain as modulated by an external factor (psychological factor). Specifically, we examined task-related functional connectivity changes underlying the differential encoding of associations with a neutral versus negative valence. The single-subject GLMs for these analyses included the general de-convolved signal extracted from the seed region, the neutral-negative task-vector and the interaction term, in addition to the six movement parameters defined earlier. The subject-specific parameter estimate for the interaction term was used as input for second-level random-effects analysis.

The seed-region was defined by the mPFC activation cluster found in the negative > neutral encoding contrast, thresholded at $p < 0.001$ uncorrected.

Statistical parametric maps were thresholded and visualized by superimposing T-contrast images onto the standard MNI-image implemented in MRIcron. Connectivity and activity were considered significant using a criterion of $p < 0.05$ Family-Wise Error cluster-corrected for the whole brain (as implemented in SPM8) after an initial voxel-level threshold of $p < 0.001$ uncorrected. Additionally, the hippocampus and amygdala were treated as regions of interest, based on their *a priori* hypothesized roles in associative memory encoding and emotional processing. Statistical inference tests about BOLD-responses in the hippocampus and amygdala were corrected for multiple comparisons in a reduced search region, defined using anatomical masks from the WFU PickAtlas Tool (version 2.4) implemented in SPM.

**Across-subject correlation**

To explore brain–behavior associations across participants, PPI measures were extracted from SPM and analyzed using SPSS. A two-tailed simple Pearson correlation and a partial Pearson correlation test controlling for test delay was performed to test the linear relation between PPI-measures on the one hand and valence-related changes in specificity on the other hand. Alpha was set at 0.05 throughout.

## Results

### Behavioral results

Subjects showed significant above-chance memory for the specific face-occupation associations (proportion correct: mean = 0.29, SD = 0.17 versus the chance level of 0.06, $t_{(101)} = 17.07$, $p < 0.001$). Memory for specific face-occupation associations was modulated by valence. Associations with negative occupational identities were less well remembered than associations with neutral occupations, indicating impaired memory for specific associations when negative (neutral correct: mean = 0.33, SD = 0.20; negative correct: mean = 0.26, SD = 0.17; difference $t_{(101)} = 4.40$, $p < 0.001$). If subjects were unable to remember the specific face-occupation associations correctly, they more often selected an occupation from the same valence category than from the other (correct valence category: mean = 0.13, SD = 0.11; valence category incorrect: mean = 0.09, SD = 0.09, difference $t_{(101)} = 4.40$, $p < 0.001$). This is indicative of a gist-like, less specific memory of the associated valence-category. This gist-like memory lacking associative specificity was more prevalent for the negatively valenced associations than for neutral ones (negative valence-only correct: mean = 0.15, SD = 0.13; neutral valence-only correct: mean = 0.11, SD = 0.12; difference $t_{(101)} = 4.26$, $p < 0.001$).

Subsequently, we estimated the proportion of trials with correctly remembered face-occupation associations relative to all trials in which at least the valence (the gist) was correct. Here, we observed a shift from specific, associative memory towards a less-specific gist-like memory for negatively valenced occupational identities (see Figure 2.2) (memory specificity neutral: mean = 0.73, SD = 0.25; memory specificity negative: mean = 0.63, SD = 0.25; difference $t_{(101)} = 4.01$, $p < 0.001$).

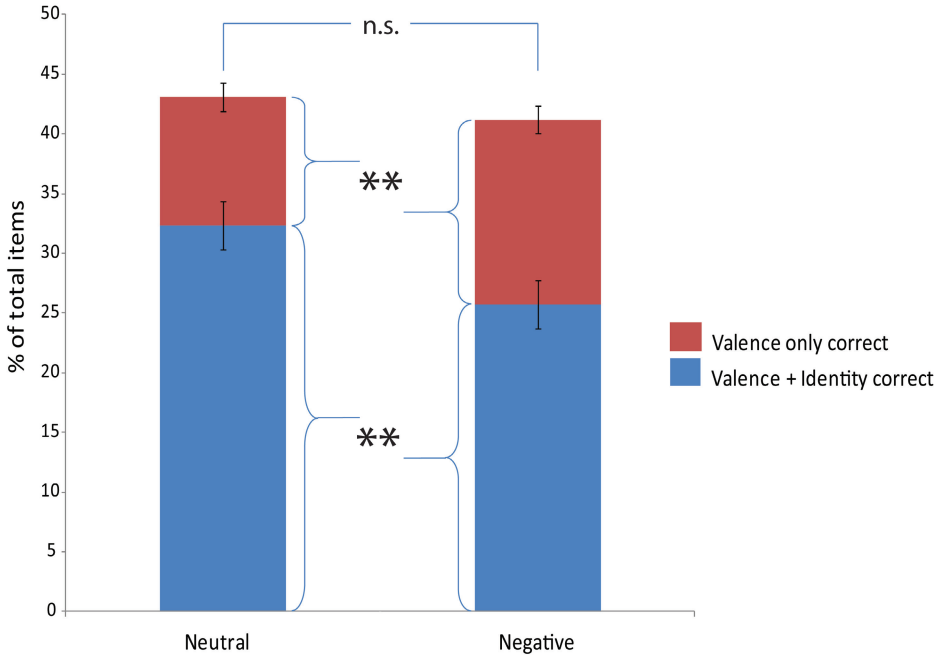



**Figure 2.2. Behavioral results.**
*The bar graph displays memory performance as proportion of total associations correctly recalled for neutral respectively negatively encoded face-occupation associations. No difference is noticeable with regards to proportion of valence-correct responses for negative and neutral associations (blue + red bars combined). A marked difference between memory recall of neutral and negative associations appears when distinguishing the proportion of those valence-correct responses where identity is also correct (blue bars) versus the proportion of valence-correct responses where identity is incorrect (red bars). Error bars reflect standard error of the mean. n.s.: not significant, **p < 0.001.*

### Imaging results

Compared to the perceptual control condition, encoding of face-occupation associations activated a set of brain regions encompassing ventral visual regions, the MTL and midline structures such as the mPFC, posterior cingulate cortex and precuneus (see Figure 2.3A). Two emotional contrasts were calculated across the face-occupation association conditions. Faces with neutral compared to negative occupational identities led to stronger activity in a large area including the left fusiform and parahippocampal gyrus and extending into the lingual gyrus and precuneus, as well as the left dorsolateral prefrontal cortex, left inferior temporal gyrus, bilateral angular gyrus, posterior cingulate cortex (cluster-level FWE corrected; $p < 0.05$), and the hippocampus bilaterally ($p < 0.05$ SVC corrected;

Figure 2.3B and Table 2.2). Conversely, increased activation for faces with negative as opposed to neutral occupations was revealed in a large cluster centering in the rostral mPFC extending into the anterior cingulate cortex (-2,52,24, cluster-level FWE corrected $p < 0.05$) (Figure 2.3c and Table 2.3).



**Table 2.2.** fMRI Results encoding neutral > negative associations.

| Brain region | Cluster size (voxels) | Cluster P | T-value | Local maxima | | |
|---|---|---|---|---|---|---|
| | | | | x | y | z |
| L fusiform / parahippocampal gyrus Bilateral precuneus | 1882 | <0.001 | 5.55 | -32 | -38 | -12 |
| L middle frontal gyrus | 468 | 0.003 | 5.55 | -24 | 12 | 54 |
| L middle occipital gyrus / L angular gyrus | 813 | <0.001 | 5.05 | -34 | -72 | 34 |
| Bilateral posterior cingulate gyrus | 859 | <0.001 | 5.04 | 8 | -34 | 38 |
| R middle occipital gyrus / R angular gyrus | 990 | <0.001 | 4.86 | 46 | -70 | 32 |
| L inferior temporal gyrus | 275 | 0.038 | 4.61 | -56 | -48 | -10 |
| L white matter | 356 | 0.013 | 4.46 | -24 | -18 | 46 |
| SVC L hippocampus | 308 | <0.001 | 5.51 | -30 | -36 | -12 |
| SVC R hippocampus | 134 | 0.006 | 4.21 | 34 | -38 | -8 |

*Clusters of voxels where activity is higher during the encoding of neutral versus negative face-identity associations. For each cluster, the local maximum is reported. Cluster p-value is whole brain corrected at the cluster level (FWE, p<0.05), or small-volume corrected (SVC) for an anatomical mask based on the AAL-atlas. All coordinates are in MNI space. L= left, R = right.*

**Table 2.3.** fMRI Results encoding negative > neutral.

| Brain region | Cluster size (voxels) | Cluster P | T-value | Local maxima | | |
|---|---|---|---|---|---|---|
| | | | | x | y | z |
| Medial prefrontal cortex | 573 | 0.001 | 4.99 | -2 | 52 | 24 |

*Clusters of voxels where activity is higher during the encoding of negative versus neutral face-identity associations. For each cluster, the local maximum is reported. Cluster p-value is whole brain corrected at the cluster level (FWE, p<0.05). All coordinates are in MNI space. L= left, R = right.*

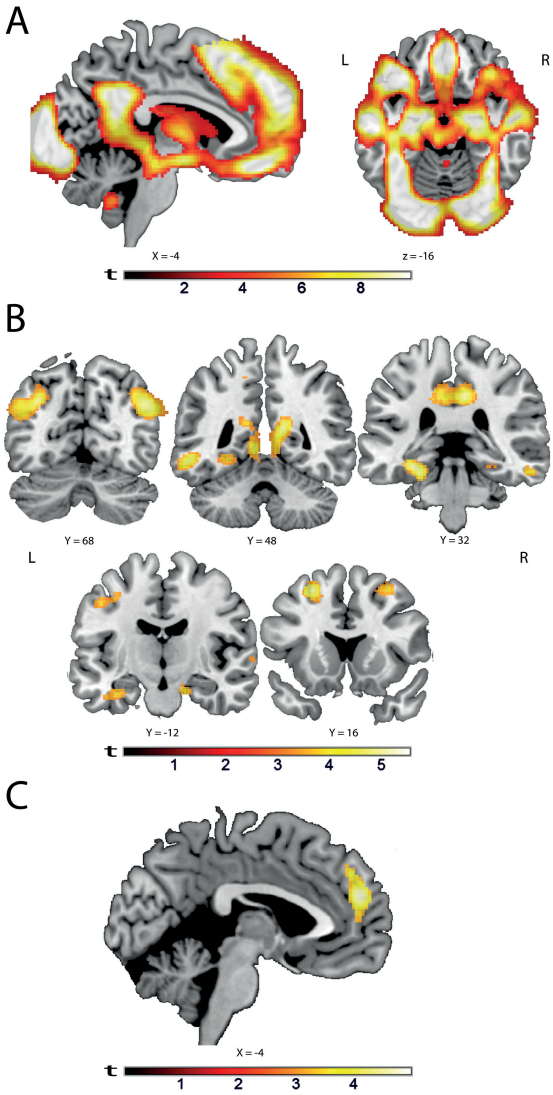



**Figure 2.3. Main activation effects of encoding faces associated with identity words.**
*Images are displayed at a threshold of p = 0.001 uncorrected, cluster size threshold: 30 voxels). A. Contrast of encoding (negative + neutral) versus the sensorimotor control task. B. Contrast of encoding neutral associations versus negative associations. C. Contrast of encoding negative versus neutral associations.*

Given our hypothesis that the mPFC is part of a network that is relevant for memory specificity and generalization, we used this mPFC cortex cluster as a seed region for a psycho-physiological interaction analysis (PPI). This enabled us to assess at which brain region is connectivity with the mPFC seed region (physiological factor) modulated by the emotional valence of the occupation (psychological factor). The mPFC displayed a relative increase in functional connectivity during negative associative encoding with a MTL cluster (FWE cluster corrected $p < 0.05$), which includes both the right hippocampus (SVC-corrected $p < 0.05$) and amygdala (SVC-corrected $p = 0.002$) (Figure 2.4A and Table 2.4). A second cluster was detected in the left inferior occipital gyrus (cluster corrected $p < 0.05$). The encoding of negative associations is thus found to be associated with a relative increase in functional connectivity between the mPFC and the MTL and early visual processing areas

**Table 2.4.** fMRI Results PPI-analysis negative>neutral.

| Brain region | Cluster size (voxels) | Cluster P | T-value | Local maxima | | |
|---|---|---|---|---|---|---|
| | | | | x | y | z |
| R hippocampus/amygdala | 176 | 0.039 | 4.91 | 28 | -10 | -12 |
| SVC R hippocampus | 43 | 0.033 | 4.59 | 26 | -12 | -12 |
| SVC R amygdala | 110 | 0.002 | 4.91 | 28 | -10 | -12 |
| L inferior occipital gyrus | 323 | 0.002 | 3.97 | -24 | -88 | -10 |

*Clusters of voxels displaying increased connectivity with the mPFC-seed during encoding of negative face-identity associations. For each cluster, the cluster local maximum is reported. Cluster p-value is whole brain corrected at the cluster level (FWE, p<0.05), or small-volume corrected (SVC) for an anatomical mask based on the AAL-atlas. All coordinates are in MNI space. L= left, R = right.*

To probe the behavioral relevance of this differential connectivity, we tested whether the valence effect on medial prefrontal–medial temporal connectivity was associated with individual differences in memory specificity. Functional connectivity between the mPFC and right hippocampus for the contrast negative versus neutral occupations correlated positively ($r_{(101)} = 0.224$, $p = 0.024$ two-tailed) with the valence related differences in memory specificity (negative memory specificity – neutral memory specificity) after removing a potent bivariate outlier (Cook's Distance > 0.25). We did not find a reliable correlation

for the right amygdala ($r_{(101)}$ = 0.068, p = 0.500 two-tailed) or the early visual cortex ($r_{(101)}$ = 0.141, $p$ = 0.161 two-tailed). When calculating a partial correlation, controlling for time delay, the correlation remains significant ($r_{(101)}$ = 0.219, p = 0.029 two-tailed).

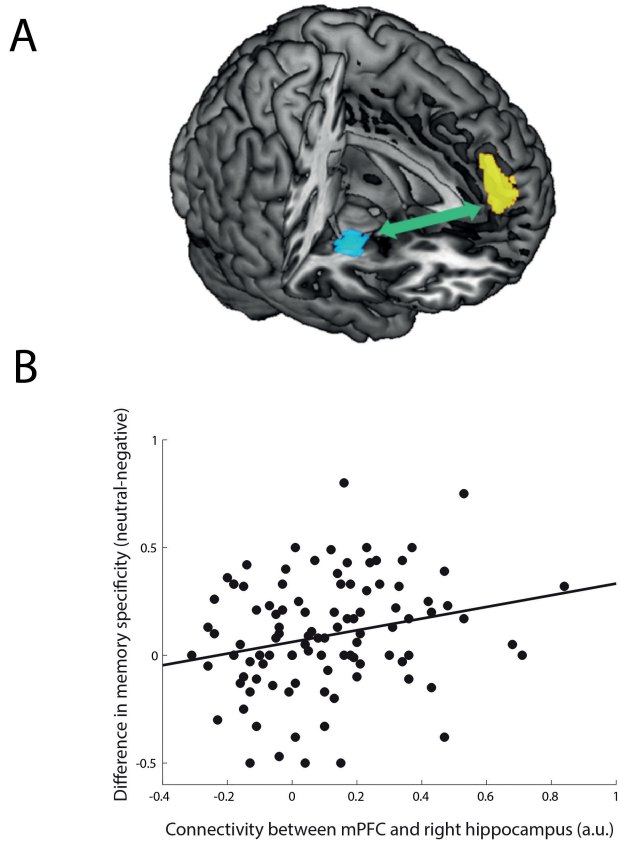



**Figure 2.4. MPFC-hippocampal connectivity.**
*A. Psychophysiological analysis with mPFC as seed region (cluster displayed in warm colors) showed a significant coactivation with the right hippocampus and amygdala (cluster displayed in cold colors), as well as the left early visual cortex (not displayed). B. The PPI-values for connectivity between the mPFC and the right hippocampus (red arrow) are extracted for across-subject correlation. A correlation between mPFC-hippocampus coupling and a loss of memory specificity for negatively valenced associations was found. The more connectivity during encoding of negatively valenced associations relative to neutral valence associations, the more the associated loss of memory specificity in the subsequent memory test ($r_{(102)}$ = -0.224, p = 0.024). a.u.: = arbitrary units.*

## Discussion

The current results build on previous literature by demonstrating a loss of memory specificity for associative information charged with a negative emotional valence. Particularly, we demonstrate that, despite an overall preserved gist memory (preserved memory for the valence of identity associated with a face), there is a loss of memory specificity evidenced by impaired retrieval of the specific associated occupational identities for the negative valence category. When encoding negative associations, a selective increase in activation is observed in the mPFC along with reduced activity in, amongst others, the bilateral hippocampus, compared to encoding neutral associations. Furthermore, we find an increase in connectivity between the mPFC and right hippocampus and amygdala during negatively valenced encoding. The behavioral relevance of this valence-specific interplay between hippocampus and mPFC is further evidenced by a negative correlation with memory specificity in an across-subject analysis. While this finding is correlational and thereby does not confirm causality, this suggests that individual differences in medial-prefrontal-hippocampal coupling may have an impact on an individual's memory specificity in an emotional context.

The behavioral finding of an emotion-induced loss of associative memory specificity is in line with both laboratory studies (Bisby & Burgess, 2014; Rimmele et al., 2011, 2012), and observations from the eyewitness literature (Christianson & Loftus, 1991; Migueles & Garcia-Bajos, 1999; Steblay, 1992) that report impairments in the recall of associated detail. The apparent paradox in the literature between an emotional enhancement effect for the recall and recognition of items, and an emotional impairment effect for retrieving associations and peripheral details can be reconciled when distinguishing between gist (general) and detailed (specific memory). The emotional enhancement effect of gist memory has been shown to be mediated by the modulatory influence of the amygdala on the MTL (Kilpatrick & Cahill, 2003; LaBar & Cabeza, 2006; McGaugh, 2004). Studies in rats have shown that the amygdala is important for gist memory, but not detailed recollection (Farovik, Place, Miller, & Eichenbaum, 2011). This relatively selective role of the amygdala in gist memory is corroborated by studies of patients with selective amygdala lesions that revealed a relative impairment of retrieving central gist-based elements, but not peripheral details of complex stimuli (Adolphs et al., 2005).

However, the neural mechanism underlying the impairment of specific detailed memories has not been sufficiently investigated yet in the context of emotional memory.

The current results suggest that interactions between the mPFC and the hippocampus may play a critical role in driving inter-individual differences in memory specificity. These results converge with recent rodent work that demonstrated the role of a circuit including the mPFC, thalamic nucleus reuniens and hippocampus in regulating the specificity of the encoding of fear memories (Xu & Südhof, 2013). On the one hand, increased activation of bilateral hippocampal and parahippocampal regions during encoding of the more specifically remembered neutral associations suggests that MTL activity is beneficial for memory specificity. On the other hand, increased mPFC connectivity with the right MTL during encoding of negative associations might indicate that the mPFC regulates memory specificity by altering encoding activity in the MTL. Based on rodent work, it could be speculated that the mPFC connects via the nucleus reuniens to inhibitory interneurons in the hippocampus (Dolleman-Van der Weel, Lopes da Silva, & Witter, 1997), effectively causing a downregulation of hippocampal processing during negative encoding and resulting in increased activity in the neutral condition. Moreover, and potentially in line with this interpretation, the activity and connectivity contrasts revealed different parts of the MTL. The mPFC seems to show an interaction with an anterior part of the right MTL including the amygdala and anterior aspects of the hippocampus. On the other hand the main effects of neutral versus negative encoding were found in more posterior regions of the hippocampus and the parahippocampal gyrus. The anterior region of the hippocampus might be particularly suited for processing more general gist-based information (Gutchess & Schacter, 2012), whereas more posterior regions might be more specialized to processing detailed information specific to an episode (Poppenk & Moscovitch, 2011). This is supported by recent investigations of hippocampal activations along the posterior-anterior axis showing that its most anterior aspect is particularly involved in encoding associative information at a coarser, global spatiotemporal scale, whereas the middle and posterior hippocampus are particularly involved in encoding associative information specific to a particular episode (Collin, Milivojevic, & Doeller, 2015; Poppenk,

Evensmoen, Moscovitch, & Nadel, 2013; Poppenk & Moscovitch, 2011). One of the mechanisms by which the connected circuit of the mPFC and hippocampus could regulate associative memory encoding is by shifting between gist and specific encoding. However, as gist-based memory was not directly measured here, future studies should look into this mechanism in more detail.

The mPFC might regulate associative memory encoding by regions in the MTL. A recent model from our group proposes that the mPFC modulates encoding resources in mnemonic regions like the hippocampus in order to maximize efficiency of memory encoding. Particularly, when new incoming information is congruent with prior knowledge only those congruent, abstract features are extracted, integrated and consolidated in a mnemonic trace (van Kesteren et al., 2013, 2014). This role of the mPFC and its connectivity with the hippocampus in optimizing the efficiency of hippocampal encoding can be extended to the current results. The mPFC could be interpreted to shift encoding resources to the central item features (e.g. a gun) that are necessary for survival, at the expense of peripheral, detailed features (e.g., print of the t-shirt that the shooter was wearing). While the current data does not allow making inferences on causality, they converge with previous literature showing that both the amygdala and mPFC can modulate specificity of associative memory encoding. In the context of the present study it could thus be speculated that the mPFC regulates hippocampal-based associative memory processes. However, the reverse direction is also possible, namely that emotional valence causes the amygdala to bias the hippocampus to shift to a gist-based encoding mode that also recruits the mPFC. Further studies should elucidate the directionality of the shift to a less specific encoding of emotional memories.

The suggested mechanism might serve to enhance encoding efficiency such that the central gist features are encoded particularly well. The net result of this process would be a shift in focus of encoding to central gist features at the expense of peripheral associated information or less relevant detail. Future studies should measure both associative memory and recognition and recall of singular items (faces and identity words) and relate these processes separately to their neural correlates. An enhanced storage of central gist features that are emotionally salient is evolutionary valuable in predicting potential threats in the

future that are globally similar to the initial experience. The loss of specificity for remembering associated detail seems less advantageous for survival. However, a shift from high specificity to higher sensitivity has potential survival value. In adverse situations it is adaptive to minimize the risk for false negatives in the detection of potential threats (van Marle, Hermans, Qin, & Fernández, 2009). However, it might manifest maladaptively in affective syndromes characterized by overgeneralization of emotional memories.

The balance between emotional gist memory and specific memory bears special relevance for clinical conditions like PTSD, depression and generalized social phobia. Pronounced emotional preferences of gist memory are reported in these patient groups (Foa et al., 2000; Moradi et al., 2000; Watkins et al., 1996), and they seem to focus on summarizing the gist rather than retrieving detail when recounting autobiographical events (Williams et al., 2007). Emotional flashbulb-memories and intrusive recollections in post-traumatic stress disorder are often vivid, but at the same time lacking in the quality of contextual detail (Bluck & Li, 2001; Brewin, 2007; Brewin, Huntley, & Whalley, 2012; Ehlers, Hackmann, & Michael, 2004; Neisser & Harsch, 1992; Talarico & Rubin, 2003). The loss of specificity in PTSD could underpin the characteristic phenomenon that seemingly neutral cues in 'safe' environments induce vivid and disturbing recollections (Lissek et al., 2005). This and other PTSD-symptoms associated with a loss of memory specificity are found to be related to altered functioning of the mPFC (Mahan & Ressler, 2012; Shin, Rauch, & Pitman, 2006). This raises the interesting possibility that overgeneralized fear memories in PTSD could be due to differential modulation of mnemonic representations by the mPFC. Similarly, depression is characterized by overgeneralization and reduced specificity of emotional memories (Dalgleish et al., 2007; Park, Goodyer, & Teasdale, 2002; Watkins & Teasdale, 2001; Watkins et al., 1996; Williams et al., 2007). Depression is also linked to altered functioning of the mPFC, especially its most ventral aspect which is often related to aberrant self-relevant and emotional processing (Lemogne, Delaveau, Freton, Guionnet, & Fossati, 2012). It would be important to investigate the role of medial prefrontal-hippocampal connectivity in emotional associative encoding further, and particularly explore their bearing on individual differences in emotional memories in clinical conditions.

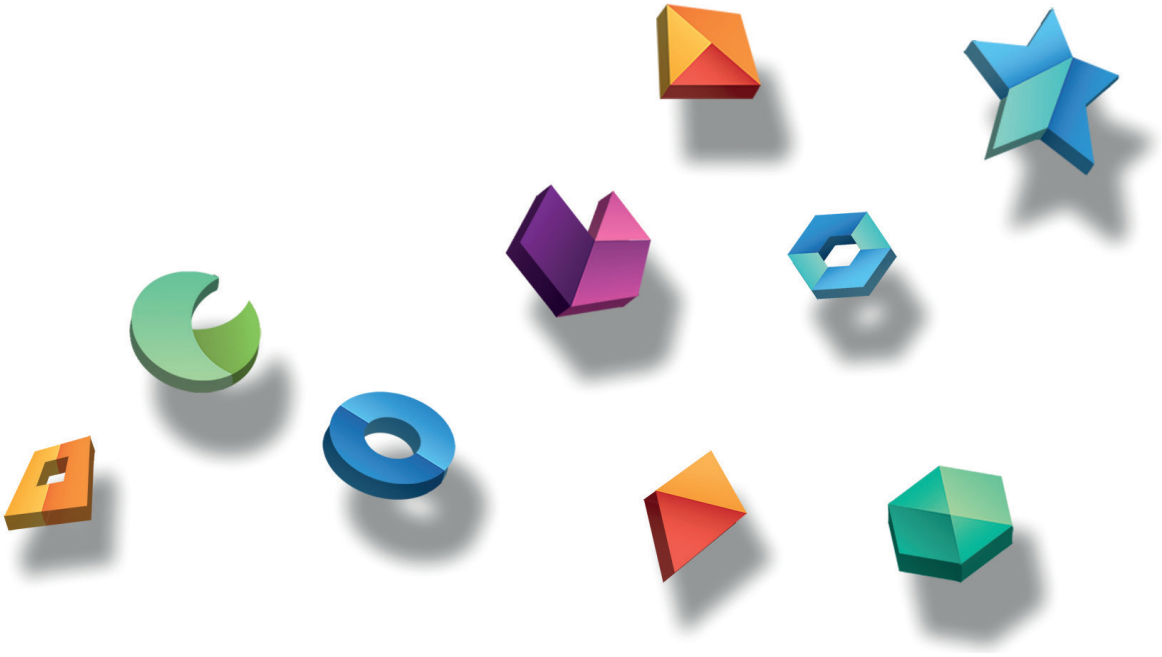
The current paradigm asked participants to study occupational identities paired with faces, which is a social task. From an evolutionary perspective, it may make sense to broadly categorize our peers into negative, neutral or positively labeled peers. This information is necessary to inform hierarchical ranking of peers and guiding our behavior towards them (e.g., avoiding 'negative' peers). However, previous studies have shown that social learning is a unique type of learning that is subserved by the amygdala (Kumaran, Melo, & Duzel, 2012). Future studies employing non-social emotional associative learning paradigms should elucidate whether the neural mechanism alluded to in this paper can be characterized as a domain-general emotional associative memory process rather than a specific type of social learning. Alternative explanations of our results are possible as well. A selective increase in mPFC activation is generally found when processing emotional information, particularly in the more rostral region of the mPFC that was found here (Etkin, Egner, & Kalisch, 2011). It could be that this activation is not specific to memory, but a general response to emotional information. Behaviorally, emotional information could be distracting to such an extent that participants are not directing sufficient attentional resources to encoding. The deficit in memory specificity is then simply due to distraction from the online task at hand. However, this explanation is unlikely as we find no overall difference in gist-based memory performance between the two valence categories, whereas this would be expected if less attention was paid when encoding negatively valenced associations. Furthermore, the relation of memory specificity to mPFC connectivity with the hippocampus (a region essential for associative memory) also speaks against this alternative interpretation. The reduction of memory specificity for negative associations could alternatively originate from retrieval effects. For instance, negative identities are semantically grouped more closely and are more interfering with each other upon retrieval. Furthermore, it could be that the effect is a consolidation effect (despite the short delay) as negative associative information has been shown to be become less specific across longer delays (Pierce & Kensinger, 2011). However, the relation observed here between memory specificity and activation and functional connectivity during encoding provides novel evidence that such effects can also originate at encoding.

There are some limitations of this study that should be noted. The study employed a blocked design that did not allow for trial-specific characterization of neural activity supporting memory. Indeed, an event-related design would have allowed specific trial effects to be discerned for successful and unsuccessful encoding. Here, encoding blocks contained different levels of encoding success, and transient emotional effects are difficult to separate from sustained mood effects. However, to ensure reliable assessments of an individual subjects' response, we leveraged on the increased detection power and design efficiency of a blocked design (Friston, Zarahn, Josephs, Henson, & Dale, 1999; Liu, Frank, Wong, & Buxton, 2001; O'Reilly, Woolrich, Behrens, Smith, & Johansen-Berg, 2012), as in previous studies of individual differences in memory (Dominique & Papassotiropoulos, 2006; Erk et al., 2010; Werner et al., 2009). Furthermore, an important additional test would have been to compare familiarity, recognition and recall for the faces and occupational identities in isolation across valence categories. This test might have found a preserved or even enhanced item recognition memory for negative identities as well as the faces associated with negative identities, further supporting a dissociation of gist and specific memory.

A further limitation is that positive emotional valence was not included in the task manipulation. One could speculate that the effect observed here is driven by arousal, not emotional valence per se, and therefore might be equally present for positive emotions. Furthermore, only men were included in the sample. The menstrual cycle has been shown to influence emotional modulation of memory and potential stress and cortisol effects in women (Sakaki & Mather, 2012). Therefore, the sample was kept as homogenous as possible, while still allowing the analysis of individual differences pertaining specifically to emotional memory encoding. Thus, future studies should have a symmetric valence manipulation including positive emotions in a sample that also includes women, perhaps controlling for menstrual cycle.

In summary, we demonstrated an emotional impairment effect on memory specificity of face-identity associations. We found a selective increase in mPFC-activity and mPFC-hippocampal connectivity during encoding of associations with a negative valence versus those with a neutral valence. Furthermore, we show

that the loss of specificity of associative memories was related to an increase in mPFC-hippocampal connectivity. This finding is in line with the role of a mPFC-hippocampal circuit regulating memory specificity, and suggests future directions for investigating the role of this circuit in regulating emotional associative memory specificity in individuals with 'overgeneralized' emotional memories.



# CHAPTER 3

## Transient medial prefrontal perturbation reduces false memory formation

This chapter is based on:

Berkers, R.M.W.J., van der Linden, M., de Almeida, R. F., Müller, N. C. J., Bovy, L., Dresler, M., Morris, R. G. M., & Fernandez, G.. (2017). Transient medial prefrontal perturbation reduces false memory formation. *Cortex, 88,* 42–52.

## Abstract

Knowledge extracted across previous experiences, or schemas, benefit encoding and retention of congruent information. However, they can also reduce specificity and augment memory for semantically related, but false information. A demonstration of the latter is given by the Deese-Roediger-McDermott (DRM) paradigm, where the studying of words that fit a common semantic schema are found to induce false memories for words that are congruent with the given schema, but were not studied. The medial prefrontal cortex (mPFC) has been ascribed the function of leveraging prior knowledge to influence encoding and retrieval, based on imaging and patient studies. Here, we used transcranial magnetic stimulation to transiently perturb ongoing mPFC processing immediately before participants performed the DRM-task. We observed the predicted reduction in false recall of critical lures after mPFC perturbation, compared to two control groups, whereas veridical recall and recognition memory performance remained similar across groups. These data provide initial causal evidence for a role of the mPFC in biasing the assimilation of new memories and their consolidation as a function of prior knowledge.

## Introduction

Knowledge acquired across prior experiences, incorporated into schemas, can exert a strong influence on the processing of incoming information. Schemas provide an internal structure or scaffold for the assimilation of congruent information, thereby enhancing memory for related, congruent information (Anderson, 1981; Bartlett, 1932). This enhancement occurs at the cost of memory for episodic detail, resulting in reduced memory specificity (Friedman, 1979; Goodman, 1980). As every episode is partly consistent and partly inconsistent with prior knowledge, competing schematic and episodic elements are weighed to integrate new memories with existing knowledge (Alba & Hasher, 1983). As such, expectations raised by schemas during memory formation can proactively interfere with memory, causing intrusions of schema-consistent, but false information at retrieval (Tuckey & Brewer, 2003).



A striking example of strong schema-based expectations leading to false memory intrusions is provided by the Deese-Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995). Studying lists of words that are semantically organized improves later recall compared to random lists of words (Bower, Clark, Lesgold, & Winzenz, 1969), but learning these lists can also provoke false memories for words that are congruent with the semantic schema but actually never studied. Here, the weighing of schematic memories is increased by study words ('rest', 'awake', 'pillow') that are semantically associated to a non-studied critical lure word ('sleep') that may therefore be more likely to intrude at retrieval (Roediger & McDermott, 1995; Stadler, Roediger, & McDermott, 1999).

The medial prefrontal cortex (mPFC) is proposed to play a critical role in leveraging prior knowledge to influence online encoding and retrieval. Whereas memory for specific episodes and general knowledge are thought to be supported by the medial temporal lobe and posterior neocortical areas, the mPFC and its connectivity with the medial temporal lobe supports the integration of novel memories into pre-existing networks of knowledge (Preston & Eichenbaum, 2013; van Kesteren et al., 2012). Indeed, previous animal and human studies have related medial prefrontal processing to a schema-congruent learning benefit (Liu, Grady, & Moscovitch, 2016; Tse et al., 2011; van Kesteren et al., 2013, 2014), as well as reduced memory

specificity (Berkers, Klumpers, & Fernández, 2016; Xu & Südhof, 2013). The mPFC could play a role in evaluating the overlap of new memories with prior knowledge, thereby weighting the influence of schematic and episodic memory components. Thus, when encoding a list related to a 'sleep' schema, a strong weighting of prior knowledge in the processing of incoming experiences causes proactive interference during later retrieval. As a result, schema-based expectations promote recall of related words that had not been encountered before, and result in faulty judgments about whether or not a semantically related word had been seen previously.

The mPFC has been implicated in memory encoding as well as retrieval after long-term consolidation (Takashima et al., 2006), but lesions to this region by themselves often do not lead to severe global impairments in declarative memory, in contrast with lesions to the medial temporal lobe (Melo, Winocur, & Moscovitch, 1999; Scoville & Milner, 1957). Specifically, damage to mPFC is reported to lead to an increase in confabulation (Schnider, 2003), and a reduction of the influence of schema on recognition memory (Spalding, Jones, Duff, Tranel, & Warren, 2015). In the DRM task, lesions situated in the mPFC covering Brodmann areas 32, 12 and 10 have been related to a reduction in false recall, whereas veridical recall remained unaffected (Warren et al., 2014). These results indicate that the mPFC contributes to false memory formation and retrieval. Lesion studies, however, have the caveat that they can be confounded by brain plasticity following damage, leading to a modification of brain connectivity, reduced spatial specificity due to damage or disruption of additional brain regions, and reduced cognitive specificity due to comorbid global impairments in cognitive functioning. Imaging studies have linked activity during encoding to subsequent false recognition only, but not false recall, and recall and recognition might involve different processes (Staresina & Davachi, 2006). Furthermore, these studies have linked encoding activity to subsequent false recognition in regions across the lateral, dorsal and ventromedial prefrontal cortex (Cabeza, Rao, Wagner, Mayer, & Schacter, 2001; Kim & Cabeza, 2007; McDermott, Gilmore, Nelson, Watson, & Ojemann, 2016).

Here, we sought to investigate further the possible role of the anterior part of the mPFC in the false memory effect in a non-clinical setting using a temporary experimental perturbation of mPFC processing that obviates the confounding

effect arising from neural plasticity. We used offline transcranial magnetic stimulation (TMS) to transiently perturb mPFC processing before participants performed the DRM-task. Previous studies have shown that by stimulating the dorsolateral prefrontal cortex it is possible to manipulate memory encoding of both verbal (Javadi, Cheng, & Walsh, 2012; Javadi & Walsh, 2012; Sandrini, Cappa, Rossi, Rossini, & Miniussi, 2001) and non-verbal material (C. M. Epstein, Sekino, Yamaguchi, Kamiya, & Ueno, 2002; Floel et al., 2004). However, to our knowledge, there has not been any brain stimulation study reported to date investigating the role of the mPFC in memory formation. Following studies that have perturbed deeper regions in the mPFC using an angled, figure-of-eight coil (Klucharev, Munneke, Smidts, & Fernández, 2011; Rushworth, Hadland, Paus, & Sipila, 2002), we aimed to perturb processing in a more anterior region of the mPFC (Brodmann area 10) with a continuous theta-burst stimulation (cTBS) protocol. This repetitive TMS protocol has been shown to induce a transient decrease in neural excitability for up to an hour (Huang et al., 2005; Huang, Rothwell, Chen, Lu, & Chuang, 2011). Our focus was on the possibility of altering the probability of 'false memories' induced by the DRM protocol. Following mPFC perturbation or a control stimulation procedure, participants encoded and recalled DRM-lists, followed by a recognition test. Based on prior literature, we predicted that cTBS directed at the mPFC would result in a reduction in false recall and/or recognition of critical lures compared to a behavioral and stimulation control.



## Materials and methods

### Participants

The present study was part of a larger study with multiple tasks. For one task not reported here there was a gender-specific hypothesis, the results of which will be reported elsewhere. Therefore, all participants included were female. Fifty-one participants were pre-screened for contraindications for participation in a TMS-study and were found to fit the criteria for inclusion (no history in participant or family of epilepsy, prior head trauma, medication, implanted neurostimulators, cochlear implant, pacemaker or intracardial wires, intake of psycho-active substances). Ten participants were excluded: one participant was excluded because we could not measure the active leg motor threshold

(see below), two participants experienced discomfort during stimulation, two participants withdrew participation, two participants were on medication, and one participants had an excessive intake of alcohol (7 units) in the 24 hours before returning for the second stimulation session, and two participants were excluded because the achieved stimulation intensity differed substantially in relation to the target stimulation intensity (less than 70% of target intensity, see Procedure Intake Session below for specifics about the method used to determine stimulation intensity). This resulted in 42 participants being included in the two stimulation groups (21 in the mPFC stimulation group and 21 in the Cz stimulation group). Furthermore, 46 age-matched participants were included in a group receiving no stimulation (behavioral control group).

Participants have been shown to be able to suppress false memories when they have knowledge about the false memory effect (McCabe & Smith, 2002). In a debriefing questionnaire administered after completing the experiment (see Appendix), participants were asked whether they knew prior to or during the task that the word lists were created as to induce false memories for words that had not been presented. Participants that answered that they knew about the purpose of the task before, or while listening to the word lists were then excluded from analysis. Overall, data from 28 subjects were excluded according to these criteria. These exclusions were unevenly distributed across participants (4 out of 21 for the group that received stimulation to the mPFC, 7 out of 21 from the group that received stimulation to the Cz, and 18 out of 46 participants in the behavioral control group), which might be due to random sampling effects since the same questionnaire and instructions were used for all participants. This resulted in a total of 59 remaining participants: 17 in the mPFC stimulation group (experimental group), 14 participants in the Cz stimulation group (TMS control group), and 28 participants in the behavioral control group. For one participant in the latter group, data for the recognition task were lost due to a technical error, but the recall data were included. As previous studies have reported a strong relation between performance on the DRM task and sleep parameters (Diekelmann, Born, & Wagner, 2010; Payne et al., 2009), data were also obtained on 'Sleep Quality' in the night before the experimental session, and the current state of 'Drowsiness' and 'Feeling Rested' (see Table 3.1 for specifics of the three groups). All participants received

payment or course credits for their contribution. All participants had normal or corrected-to-normal vision and provided written informed consent according to the guidelines of the local ethics committee (Commissie Mensgebonden Onderzoek region Arnhem-Nijmegen, The Netherlands).

**Table 3.1.** Participant sample characteristics.

| | Behavioral control N=28 | MPFC stimulation N=17 | Cz stimulation N=14 | Group differences |
|---|---|---|---|---|
| Age (yrs) | 21.71 (0.53) | 21.59 (0.57) | 23.29 (0.65) | 5.219 (0.07) |
| Sleep Quality | 5.86 (0.25) | 5.29 (0.19) | 5.00 (0.23) | 1.138 (0.57) |
| Drowsiness | 2.96 (0.24) | 2.94 (0.35) | 2.86 (0.40) | 0.070 (0.97) |
| Feeling Rested | 4.39 (0.22) | 4.71 (0.17) | 4.64 (0.33) | 0.915 (0.63) |
| AMT hand (%) | N.A. | 26.94 (1.09) | 25.07 (1.17) | 1.587 (0.21) |
| AMT leg (%) | N.A. | 36.88 (1.53) | 35.56 (2.27) | 1.201 (0.27) |
| TMS Intensity (%) | N.A. | 28.35 (1.10) | 26.43 (1.42) | 1.563 (0.22) |

*Information on age, stimulation parameters and sleep characteristics on the night prior to the experimental session for the three groups. All subjects were female. Descriptive statistics are reported as means with standard error of mean (S.E.M.) in brackets. Sleep Quality, Feeling Rested and Drowsiness were rated by participants on a six-point scale. Stimulation parameters are reported as a percentage of maximum stimulation output. The last column reports $\chi^2$-statistics on group differences (p-value between brackets).*

### Intake session procedure

TMS protocols were applied using a biphasic pulse configuration using a MagVenture MC-B70 Butterfly coil connected to a MagPro-X100 stimulator (MagVenture, Farum, Denmark). This coil uses two angled windings (inner diameter: 27mm, outer diameter 97mm) to increase the effectiveness of stimulating relatively deep brain areas. Stimulation intensity was defined by measuring active motor thresholds during the intake session. Specifically, the toe/leg representation in the primary motor cortex (Cz) is located at a comparable depth level in the interhemispheric wall compared to the target location in the mPFC (see also Klucharev et al., 2011). Therefore, stimulation intensity was defined as 80% of the measured active leg motor threshold. Further, to ensure that stimulation intensity was within established safety guidelines (Rossi, Hallett, Rossini, & Pascual-Leone, 2009), the active motor threshold was also obtained

for the right hand as a reference. As such, if 80% of the active leg motor threshold exceeded 120% of the active hand motor threshold, the intensity was adjusted as to fall below 120% of the active hand motor threshold (in each stimulation group one participant was excluded where the achieved stimulation intensity with this procedure was lower than 70% of target intensity, see Participants). Electrodes were placed on both the first dorsal interosseous (FDI) muscle of the right hand (belly-tendon montage), and the tibialis anterior (TA) muscle of the right leg (placed collaterally at approximately 1/3 of the length of the muscle from the top). The electromyography-signal was measured at a 1 kHz sampling rate and bandpass filtered at 1-1000 Hz using an EKIDA DC amplifier (Ekida GmbH). During stimulation of the lateral hand area of the primary motor cortex, the active motor threshold was established as the minimum stimulation intensity of single pulses that produced a liminal EMG response in 50% of trials during isometric contraction of the FDI muscle. The active motor threshold was similarly defined for the midline toe/leg area of primary motor cortex (Cz) during isometric contraction of the TA muscle.

Coil position for the repetitive stimulation was established by drawing locations using the 10-20 system on a tightly-worn swim cap. The target stimulation was delivered to target the mPFC, whereas control stimulation was delivered to target the midline toe/leg representation of the primary motor cortex (Cz). First, the Cz was outlined with a removable marker as half the distance from left to right tragus, as well as the distance from nasion to inion. Second, the mPFC stimulation site was defined as two-thirds the distance from vertex to nasion (see Figure 3.1). Therefore, the coil was positioned tangentially to the skull, and oriented with a 90° angle with reference to the sagittal midline, with the handle either pointing to left-lateral or right-lateral direction (counterbalanced across subjects). This orientation is optimal for inducing an electrical field in the lateral-medial direction, perpendicular to the cortical layers in the medial wall of the prefrontal cortex (Laakso, Hirata, & Ugawa, 2013). The stimulation protocol consisted of a 40s continuous theta-burst train (Huang et al., 2005), which consists of bursts of three pulses administered at a rate of 50 Hz, that are themselves repeated at a rate of 5 Hz. The stimulation protocol was administered over either mPFC or Cz location (see Table 3.1 for stimulation parameters of the two stimulation groups).

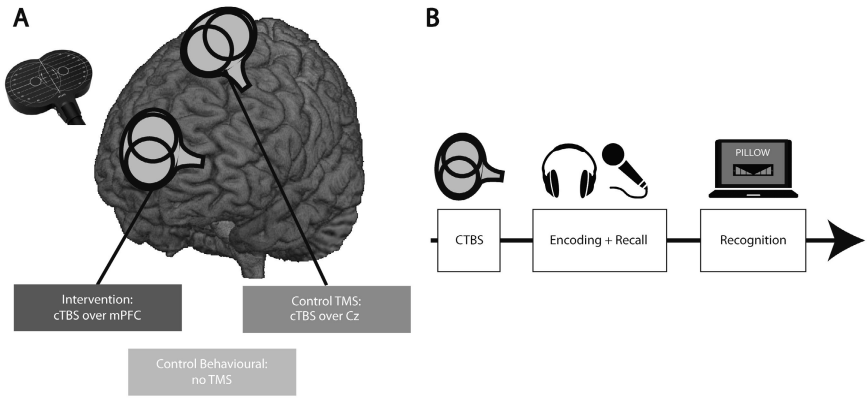





**Figure 3.1. Experimental procedures.**
*A. The three groups consisted of an intervention group receiving stimulation over mPFC, a stimulation control group receiving stimulation over Cz, and a behavioral control group. B. The experimental session started with administration of cTBS using TMS. Next, participants listened to and immediately recalled the word lists. Lastly, they performed a recognition task on a computer.*

### DRM task

In the DRM task, participants are presented with lists of 15 semantically associated words that are all related within each list to an omitted critical lure word. At free recall and recognition, these critical lures are then often falsely recalled and recognized (critical intrusions) as having been part of the studied word list (Roediger & McDermott, 1995). Twenty-two English DRM lists were selected based on reported norms from a larger set of 36 lists (Stadler et al., 1999). Specifically, lists were selected that in at least 30% of all participants had evoked false recall of the critical lures. The lists were translated into Dutch. An initial pilot study (N=6) was performed to exclude lists that did not produce any false recall of critical lures. The lists were read out by the experimenter at a rate of one word per second, followed immediately by a prompt for the recall of the list words. Four lists were rejected that did not prompt false recall of critical lures in the pilot subjects. Next, the resulting eighteen lists were read out by the experimenter at a speed of one word per second and recorded using the free audio editing program Audacity (http://audacity.sourceforge.net). From each list, the first, eighth and tenth words were included for the recognition test (54 old items), intermixed with a further 18 critical lures and 54 semantically unrelated lures taken from the discarded DRM lists.

**Experimental session procedure**

The stimulation protocol was administered while the participants remained seated in a comfortable chair in the stimulation lab. Next, participants were relocated to a behavioral lab, explained the purpose of the task through verbal and written instructions and given a headset to wear (mean delay between TMS and task: 349.09s, S.E.=11.59s). The audio files for each of the eighteen lists were played in randomized order to the participants through a headset. After each list, the participants were instructed to verbally recall as many words as possible from the list. Recall was self-paced, there was no time limit, and participants pressed a button to hear the next list when they thought they would not be able to recall any more words from the list. Recall was scored by the experimenter present in the same room, but also recorded with a microphone throughout, and scored by a secondary rater who was blind to group membership of participants.

After completion of recall for all eighteen lists, instructions for the recognition task were given and the task was started (see Figure 3.1). In the recognition task, the 54 old items, 54 unrelated lures and 18 critical lures were presented in randomized order. Each word was presented on screen in white font against a gray background, with a rating scale containing six answer options and their corresponding keyboard buttons displayed ([A]: 'certainly old', [S]: 'probably old', [D]: 'perhaps old', [J]: 'maybe new', [K]: 'probably new', [L]: 'certainly new'). This task was also self-paced, and after pressing a button the response was visible for a further two seconds. Each trial was then separated by a three second interval. Lastly, the participant completed the debriefing questionnaires.

## Analysis

Recall performance was scored manually by the experimenter during the session. An independent rater who was blind to group membership also rated recall scores by listening to the audio recordings for a subset of participants (N=41, including all participants in stimulation conditions except two where audio files failed to record). Inter-rater reliability was assessed using interclass correlation (ICC; Howell, 2012) and was found to be consistent for recall of studied items (ICC=0.740, $p<0.001$) and recall of critical intrusions (ICC=0.969, $p<0.001$). The experimenter's scores were used with confidence hereafter. Recall was defined as mean proportion

recalled of the total amount: studied items as words recalled compared to all items presented, and critical items as critical intrusions recalled compared to all critical lures. Recognition performance was characterized using d'. At low confidence levels, there were significant differences between the amount of 'old' and 'new' responses for lures and critical lures. Therefore, responses were collapsed across all confidence levels (high = 'certainly' old or new, middle = 'probably' and low = 'maybe'). The corrected recognition rate used the proportion of hits from all studied items. Similarly, corrected critical endorsement rate used the proportion of critical items endorsed versus all critical items, and the false alarm rate used the proportion of unrelated lures endorsed versus all unrelated lures. Rates of 1 were replaced with 1-[1/(2 * N)] before calculating d', with N equal to the total number of trials (Macmillan & Creelman, 2005). Qualitative sleep ratings on the six point Likert scale were converted to a quantitative scale from 1 till 6. Statistical analyses were performed using IBM SPSS statistics (version 24). General linear models were used to evaluate the effect of group on false recall and correct recall respectively, while controlling for qualitative sleep parameters. If significant group effects were found, follow-up comparisons were performed on the estimated marginal means using post hoc independent samples t-tests (two-tailed). Due to non-normal data distributions, the nonparametric Kruskal-Wallis H was used to test for group differences on all other variables.



## Results

### Recall

Participants freely recalled on average more than half of all studied words (mean proportion correct: 0.57, S.E.=0.01), and falsely recalled over a third of all critical intrusions (mean proportion: 0.36, S.E.=0.03). 'Sleep Quality' as rated by participants for the night before the experimental session correlated positively with false recall ($\rho(59)=0.320$, $p=0.013$), whereas the rated 'Drowsiness' correlated negatively ($\rho(59)=-0.266$, $p=0.042$) with correct recall. There were, however, no significant overall group differences in these sleep parameters (see Table 3.2).

There were no significant effects found of stimulation group on true recall performance ($F(2, 53)=0.379$, $p=0.687$, Partial Eta Squared=0.014), while

controlling for sleep parameters. There was a significant effect of stimulation group ($F(2, 53)=3.271$, $p=0.046$, Partial Eta Squared=0.110) on the recall of critical intrusions, while controlling for sleep parameters. The experimental group receiving TMS to the mPFC was found to recall fewer critical intrusions than the TMS control group (mean difference: 0.16, $p=0.022$) and the behavioral control group (mean difference: 0.12, $p=0.041$, while there were no differences in recall of critical intrusions between the two control groups (mean difference: 0.04, $p=0.55$, see Table 3.2 and Figure 3.2). Thus, stimulation of the mPFC appeared to decrease false recall of critical intrusions for DRM-lists encoded after stimulation in those subjects that were naive to the purpose of the task, and while controlling for sleep parameters.

The direction of the coil in the stimulation groups did not have an effect on true recall (mPFC stimulation group, mean difference left-right handle orientation: 0.06, $p=0.12$, Cz stimulation group, mean difference left-right: 0.02, $p=0.66$) nor false recall performance (mPFC stimulation group, mean difference left-right: 0.01, $p=0.88$, Cz stimulation group, mean difference left-right: 0.06, $p=0.66$). Furthermore, there were also no group differences in the time participants took to recall the lists ($\chi^2(2) = 1.504$, $p = 0.47$), or the serial order in which the critical intrusion was recalled ($\chi^2(2) = 1.991$, $p = 0.37$).

**Recognition**

Participants were better than chance in recognizing studied items (mean d' studied items: 1.88, S.E.=0.08, versus a chance level of d'=0, (Wilcoxon one-sample signed rank test, $Z_{(57)}=6.593$, $p<0.001$), but also more likely than chance to endorse critical lures as studied (mean d' endorsed critical intrusions: 2.19, S.E.=0.10, versus a chance level of 0, Wilcoxon one-sample signed rank test, $Z_{(57)}=6{,}616$, $p<0.001$). There were no group effects on recognition of studied items ($\chi^2(2) = 2.702$, $p = 0.26$) or critical intrusions ($\chi^2(2) = 1.576$, $p = 0.46$).

**Table 3.2.** Group performance for recall and recognition tasks.

| | | Behavioral control N = 28* | mPFC stimulation N = 17 | Cz stimulation N = 14 |
|---|---|---|---|---|
| **Recall** | Studied items total | 0.57 (0.02) | 0.55 (0.02) | 0.58 (0.02) |
| | Critical intrusions total | 0.37 (0.04) | 0.29 (0.04) | 0.42 (0.05) |
| | Time to complete (s) | 1108.10 (63.15) | 1052.75 (51.03) | 1169.26 (67.57) |
| | Serial position (#)** | 6.57 (0.32) | 6.70 (0.34) | 7.21 (0.38) |
| **Recognition** | Hits | 0.79 (0.02) | 0.77 (0.02) | 0.79 (0.03) |
| | False alarms | 0.18 (0.04) | 0.17 (0.02) | 0.14 (0.03) |
| | Critical intrusions | 0.84 (0.03) | 0.81 (0.05) | 0.88 (0.03) |
| | D' studied items | 1.83 (0.14) | 1.78 (0.12) | 2.08 (0.12) |
| | D' critical intrusions | 2.13 (0.14) | 2.07 (0.19) | 2.43 (0.19) |

*Descriptive statistics of proportions and d' are reported as means with S.E.M. in brackets. *Data for the recognition data was missing for one participant. **Data on the serial position of the recall of the critical intrusion was not included for some participants due to missing audio-recordings (included data for Behavioral group: N = 24; Cz group: N = 13, mPFC group: N = 16).*

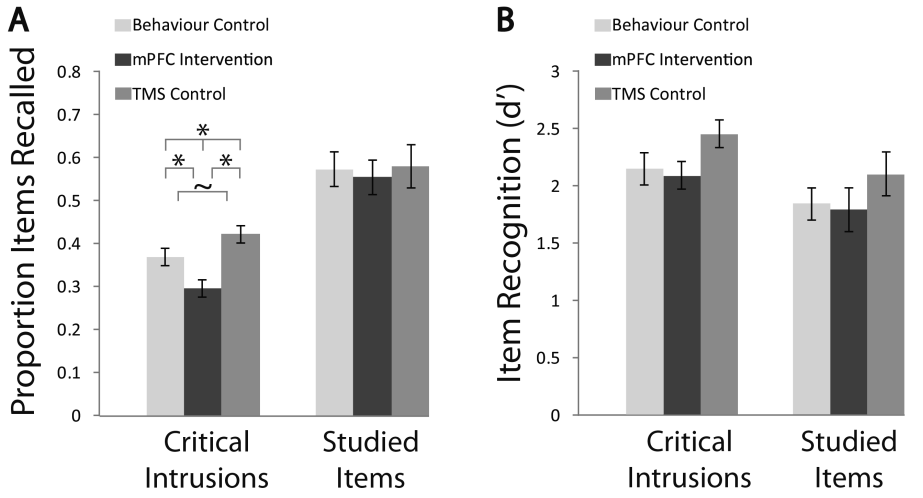



**Figure 3.2. Performance on the DRM task.**

*A. Free recall performance on critical intrusions and studied items for the intervention group and two control groups. B. Overall recognition performance on critical intrusions and studied items for the intervention group and two control groups. Error bars indicate S.E.M.; *p < 0.05; ~p > 0.1.*

## Discussion

This study set out to test the hypothesis that the mPFC is involved in evaluating the congruency of new memories with prior knowledge, thereby weighting the influence of schematic and episodic memory components. This hypothesis predicts that when encoding a list of semantically related words in the DRM paradigm ('rest', 'awake', 'pillow'), the weighting of schematic memory components in memory inferences is increased along with the probability of falsely recalling non-studied words of the same schematic category (i.e. the non-studied critical lure - 'sleep'). Experimental perturbation of the brain region presumed to support this process, the medial prefrontal cortex (mPFC), should therefore induce a decrease in false memory formation. The results are that the application of an inhibitory TMS protocol to the mPFC before studying DRM lists indeed induced a decrease in false recall of critical lures for the lists studied after stimulation, whereas recognition and veridical recall performance appeared unaffected compared to control groups. This pattern of results is in line with previous patient studies (Warren et al., 2014), and provides initial causal evidence for a specific role of the mPFC at the intersection of prior knowledge, memory formation, and memory specificity.

Certain procedural features and limitations of the study should be noted. First, the use of temporary interruption of the mPFC using a *c*TBS protocol applied via TMS obviates issues associated with brain plasticity that can arise with permanent lesions and the study of patients with frontal lobe damage. However, this approach rests on the assumption that *c*TBS has inhibitory effects on synaptic plasticity lasting up to an hour in the medial wall of the prefrontal cortex (Hayward et al., 2007; Huang et al., 2005). Particular uncertainty surrounds the depth at which stimulation had an effect on cortical excitability. Of the lesion extent reported in Warren et al. (2014), it is unlikely that BA regions 32 and 12 were targeted, but plausible that BA region 10 in the anterior medial prefrontal cortex was reached, potentially explaining the behavioral effects reported here. Further simulation studies should formally model physiological effects of stimulation using this specific coil and stimulation site and empirically verify these with functional neuroimaging studies in order to optimize the current design further.

Second, because encoding of the word lists was immediately followed by recall, it is not possible based on the current study to differentiate between encoding or recall processes. Future studies investigating the role of the mPFC in schema-based false memory inferences should temporally separate encoding and recall, and pinpoint effects on list encoding and recall separately. Third, we did not find any group effects on false recognition, most likely due to ceiling effects that are also reported previously (Warren et al., 2014). Fourth, the *c*TBS was applied to the mPFC in the experimental group, with the coil placed in close proximity to peripheral facial muscles, pain and sensory receptors. However, in the stimulation control group, the coil was placed relatively distant from peripheral musculature and receptors. Therefore, the *c*TBS protocol could have induced stronger muscle contractions when applied to the mPFC, with potentially stronger sensations or discomfort, pain and emotional distress in this condition. These effects could confound task effects, although they would likely induce a general recall impairment rather than the specific impairments reported here. Muscle contractions were observed to be limited, and over the course of the experiment only six participants from the two stimulation groups reported some minor experiences of discomfort, pain or stimulation of eyes and ears (four following mPFC stimulation, two following Cz stimulation). This relatively low level of discomfort compared to lateral prefrontal stimulation could be due to the fact that the medial facial muscles are smaller than lateral facial muscles.

Third, recall performance was found to display a relationship with self-rated sleep parameters. Previous studies have found a relationship between sleep and recall performance on the DRM-task, but these looked mostly at the influence of post-encoding sleep on subsequent false memory performance (Diekelmann et al., 2010; Payne et al., 2009). For instance, one study found that sleep (versus comparable periods of wakefulness) promoted both correct and false word recall in general, and selectively preserved false memories across increasing encoding-test delays (McKeon, Pace-Schott, & Spencer, 2012; Payne et al., 2009). Another study found that both sleep and sleep deprivation in the period between encoding and retrieval, promoted false recall of critical lures (Diekelmann et al., 2010). On recognition tasks, sleep has been found to promote correct recognition performance, but also false recognition of critical lure items (Darsaud et al., 2011). Paradoxically,

one study found that sleep deprivation, but not sleep, promoted false recognition of critical lure items. Thus, sleep and paradoxically also sleep deprivation, may promote false memory performance on the DRM-task. None of these studies has, as of yet, probed the influence of sleep prior to both the encoding and recall of the word lists. We report here that self-rated sleep quality promotes false recall of critical lure words. It could be that better sleep prior to memory encoding might make the brain more likely to shift to an encoding mode where it draws more on prior knowledge to encode incoming information, and thereby results in an incorrect inference that the critical lure was in, in fact, presented. On the other hand, self-rated drowsiness unsurprisingly displayed a negative relationship to correct recall performance, which may have been due to a lack of cognitive control over recall (Badre & Wagner, 2002; Simons & Spiers, 2003). These findings highlight the need to more fundamentally investigate the relationship between pre-encoding sleep, encoding and subsequent retrieval. Moreover, they reveal the importance of taking sleep into account when investigating memory in general, and in particular false memory performance, memory inferences and distortions.

The results reported here build and expand on previous studies that highlight a role of the mPFC in integrating novel experiences with existing schematic knowledge. The extent of overlap between novel experiences and schema determines the manner in which new experiences are encoded. If the overlap between schema and novel experiences is high, mnemonic processing relies more on schematic components rather than episodic memory components. Studies in animals (Tse et al., 2011), and human lesion (Spalding et al., 2015) and functional imaging studies (Bein, Reggev, & Maril, 2014; Liu et al., 2016; van Kesteren et al., 2013, 2014) have related processing in the mPFC to a learning benefit for information that is congruent with prior knowledge. However, the exact localization of these effects within the mPFC remains unprecise, as imaging studies have implicated both ventral (van Kesteren et al., 2013, 2014) and rostromedial aspects of the mPFC (Bein et al., 2014; Brod, Lindenberger, Wagner, & Shing, 2016). False memories have been linked to the mPFC in human lesion studies (Warren et al., 2014) and functional imaging studies, although precise localization of these effects in the medial prefrontal cortex is also lacking (Cabeza et al., 2001; Kim & Cabeza, 2007; McDermott et al., 2016).

Anatomical evidence from rodents and non-human primates indicate that the mPFC is connected, through the thalamus, with the medial temporal lobe and posterior representational areas (Aggleton & Brown, 1999; Aggleton, Desimone, & Mishkin, 1986; Amaral & Cowan, 1980; Cassel et al., 2013; DeVito, 1980; Hoover & Vertes, 2012; Hsu & Price, 2007; Van Der Werf, Jolles, Witter, & Uylings, 2003; Vertes, Hoover, Do Valle, Sherman, & Rodriguez, 2006; Vertes et al., 2007). Recently, Xu and Südhof demonstrated in mice that the anatomical circuit including the hippocampus, nucleus reuniens of the thalamus and the mPFC determines the specificity of memory encoding (Xu & Südhof, 2013). Functional imaging studies in humans have also shown that the mPFC is functionally connected to regions in the medial temporal lobe and temporo-parietal association cortices (Andrews-Hanna et al., 2010; Greicius, Krasnow, Reiss, & Menon, 2003; Öngür & Price, 2000; Saleem et al., 2008; Uddin et al., 2010), potentially also mediated by the thalamus (Thielen et al., 2015). Therefore, this brain region can function as a higher-order convergence zone of perceptual and mnemonic information. The anterior aspect of the mPFC, which was likely targeted here, has specifically been shown to connect both anatomically (Liu et al., 2013) and functionally to the medial temporal lobes in various task contexts (Bein, Reggev, & Maril, 2014; Benoit, Gilbert, & Burgess, 2011; Berkers et al., 2016; Brod, Lindenberger, Wagner, & Shing, 2016; Liu et al., 2016). In such a functional network, the medial temporal lobe is typically implicated in episodic memory (Nyberg, Mclntosh, Houle, Nilsson, & Tulving, 1996; Schacter & Wagner, 1999), whereas the angular gyrus and temporal regions are involved in storing higher-order conceptual knowledge (Bonner, Peelle, Cook, & Grossman, 2013; Bonnici, Richter, Yazar, & Simons, 2016; Price, Bonner, Peelle, & Grossman, 2015; Wagner et al., 2015). The mPFC, in turn, has been found to be involved in leveraging prior knowledge for making novel inferences (Zeithamova et al., 2012), simulating the future (Benoit et al., 2011; Benoit, Szpunar, & Schacter, 2014) and guiding decision-making (Bechara, Damasio, Tranel, & Damasio, 1997; Venkatraman, Rosati, Taren, & Huettel, 2009; Wang et al., 2014). Therefore, the mPFC might be ideally positioned to weigh the relevance of schematic elements for the current situation, mediating concurrent benefits in schema-congruent learning and schema-based memory inferences. In the context of a list of semantically associated words, schematic elements are weighed more heavily compared to detailed episodic components, at the same time potentially promoting recall

(Bower et al., 1969) and false schema-based memory inferences (Roediger & McDermott, 1995).

According to this account, perturbation of the mPFC should not only decrease the propensity for making false schema-based memory inferences, but also impair recall performance by abolishing schema-benefits for veridical recall. Here, we found no effects on veridical recall, nor were they found in patients with mPFC lesions (Warren et al., 2014). This lack of impaired veridical memory could be explained by the construction of DRM lists, which is done by selecting the strongest semantic associates of specific critical lure words (Roediger & McDermott, 1995; Stadler et al., 1999). Therefore, the study words are not selected on basis of the strength of their mutual semantic associations, although they are associated indirectly by virtue of the connecting critical lure word. For instance, the word list with the critical lure word 'black', contains study words with a strong semantic link to the critical lure word such as: 'white', 'coal', 'night' and 'funeral'. However, here the semantic link among the study words themselves is only indirect. As such, false recall of critical lures is strongly affected by schema memory components, whereas veridical recall is relatively more dependent on episodic memory. In such a scenario, lesions or experimental perturbation of the mPFC should indeed more strongly affect false memory for critical lures rather than veridical memory of studied words.

The stimulation protocol used is beneficial for stimulating deeper brain regions, but comes with the necessary side-effect of stimulating regions nearer the convexity. For instance, the frontal pole is likely affected by the protocol, which has been implicated in future-thinking and prospective memory (Burgess, Scott, & Frith, 2003; Okuda et al., 2003), and encoding future goals (Tsujimoto, Genovesio, & Wise, 2011). Indeed, the encoding task likely involved a prospective component, as subjects knew at encoding that they would be tested later on the word lists. However, if a down-regulation of prospective future retrieval goals were the primary mechanism driving the selective effects on false memory recall found here, similar deficits would have been found on veridical recall. Furthermore, due to the increased diameter of the coil windings, stimulation probably also affected dorsolateral prefrontal cortices, which have been widely implicated in the top-

down deployment of attention (Buschman & Miller, 2007) and working memory maintenance (Curtis & D'Esposito, 2003). One could speculate that working memory maintenance was required to keep words online when encoding the word lists, and it is only when several words are kept online simultaneously that the common theme of the list can be detected. In such a case, down-regulation of working memory maintenance would prevent participants from detecting the common theme, which would prevent false inferences of having encoded the critical lure. Although plausible initially, it has been shown that a reduction in working memory capacities is actually associated with an increase in false memories on the DRM-task (Peters, Jelicic, Verbeek, & Merckelbach, 2007) and misinformation paradigms (Zhu et al., 2010). Furthermore, down-regulation of working memory maintenance would probably also affect veridical recall of the correct words, rendering this mechanism an unlikely explanation for the reported effects.

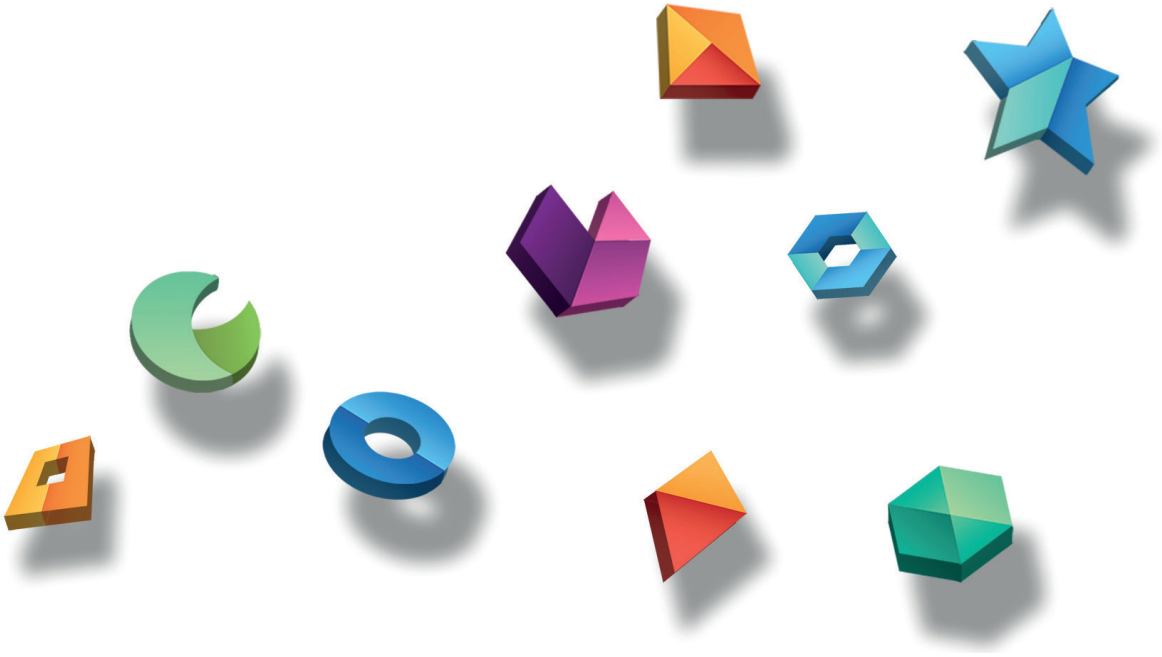


Another alternative explanation within the schema framework can be made for the reported effects. It could be that the basic deficit underlying the selective effects on false recall are explained by a decreased ability to access, activate or construct prior knowledge while encoding word lists. Whereas retrieval of semantic information is typically related to more lateral regions of the prefrontal cortex in imaging studies (Badre & Wagner, 2002; Thompson-Schill et al., 1997; Wagner, Paré-Blagoev, Clark, & Poldrack, 2001), stimulation of medial prefrontal cortex could have reduced the access to prior knowledge while encoding novel word lists. Indeed, while it can be argued that a down-regulation of the medial prefrontal cortex reduced the influence of prior knowledge on novel memory encoding, future studies should elucidate whether the critical task of the medial prefrontal cortex is to access or construct schematic themes while encoding new information, or evaluating already accessed schematic components while encoding new information.

Further work is necessary to develop methods of stimulating the medial surface of the prefrontal cortex. Such a more focused protocol would be of interest to investigate schema-based memory processing. Previous brain stimulation studies of memory have only targeted parietal cortex (Bonnì et al., 2015; Wang et al.,

2014), primary sensory cortex (Waldhauser, Braun, & Hanslmayr, 2016) or lateral prefrontal cortex (Epstein et al., 2002; Floel et al., 2004; Hanslmayr, Matuschek, & Fellner, 2014; Javadi et al., 2012; Javadi & Walsh, 2012; Sandrini et al., 2001). This study adds to these approaches by documenting a protocol for off-line stimulation of the mPFC which does not suffer from the practical limitations of earlier protocols. For instance, earlier studies targeting the mPFC have used the double cone coil (Hayward et al., 2007; Klucharev et al., 2011), which has a specific geometry that limits placement such that only midline regions can be targeted, and the lateral windings cover a significant portion of the lateral prefrontal cortex. With a more conventional, smaller figure-of-8 coil it might be possible to more specifically stimulate midline regions. This methodology should, crucially, be validated using formal modeling and measurement of physiological effects of stimulation using functional neuroimaging.

In sum, as seminal studies on false memories have shown (Loftus, 1996), our memory is not like a book that can be written and read out again in a rote manner. Rather, it is a generative, reconstructive process which relies on episodic memory components and inferences based on our prior knowledge or schemas to reconstruct prior experiences and simulate future scenarios. Memory distortions are prevalent even in people with superior autobiographical memory (Patihis et al., 2013). Therefore, it is of utmost importance to understand how these memory distortions arise and how the brain is involved in their construction. Our results provide initial evidence that an experimental stimulation intervention targeting the mPFC can decrease the influence of schema on memory formation.



# CHAPTER 4

## Neural dynamics of accumulating and updating linguistic knowledge structures

This chapter is based on:

Berkers, R.M.W.J., van der Linden, M., Neville, D.A., van Kesteren, M.T.R., Morris, R.G.M., Murre, J.M.J., & Fernández, G. (2018). Neural dynamics of accumulating and updating linguistic knowledge structures. *bioRxiv*, 495168.

## Abstract

Knowledge is acquired by generalization and integration across learning experiences, which can then be applied to future instances. This study provides novel insights into how linguistic associative generalized knowledge is acquired by systematically tracking schematic knowledge formation while participants were learning an abstract artificial language organized by higher-order associative regularities. During learning, we found activity in the left inferior frontal gyrus in response to knowledge updating during feedback presentation, as well as in response to available accumulated knowledge during retrieval. A complementary signal was found in the caudate nucleus, where activity correlated with the availability of recently acquired knowledge during retrieval, suggesting it initially supports the retrieval of knowledge. Furthermore, we found that activity in a set of regions, including the medial prefrontal cortex and hippocampus, scaled with accumulated knowledge during feedback presentation, which might be indicative of increased generalization of features of the hierarchical knowledge structure. Together, these results provide a mechanistic insight into how linguistic associative knowledge is acquired by generalization across repeated learning experiences.

## Introduction

Associative knowledge structures, or schemas, capture consistent relationships amongst low-level perceptual features as well as higher-order concepts across multiple episodes (Ghosh & Gilboa, 2014; van Kesteren et al., 2012). Humans are driven towards discovering structure across seemingly arbitrary low-level contingencies. In one demonstration, people studied geometric figures linked to arbitrary letter strings (Kirby et al., 2008). When cued with the geometric figures and asked to type the associated label, they were generally poorly reproduced. Interestingly, when using one participants' output as labels for the next participant, and repeating this process iteratively, an artificial language evolved with a higher-order hierarchical structure. This schematic structure was imposed by iterative errors that drove the language to attain compositionality. One example of such an evolved language involved a structure whereby each syllable uniquely denoted a perceptual feature: the first syllable denoting the color; the second, geometric shape; and the last syllable denoting the movement trajectory of the figure (see left panel of Figure 4.1). Moreover, the language developed a structure that was consistent across individual exemplars whilst remaining uniquely identifiable for individual exemplars.



Schematic knowledge is thought to be stored in associative neocortical structures (Bartlett, 1932; Ghosh & Gilboa, 2014; van der Linden et al., 2017; van Kesteren et al., 2012; Wagner et al., 2015), consisting of categorical and hierarchical nodes (Barsalou, 2009; Miller, Freedman, & Wallis, 2002). Schematic knowledge is acquired by extracting regularities across episodes to build a structure of higher-order relationships that can then be applied to novel instances (Brady & Oliva, 2008). Often, activated prior knowledge constrains the acquisition of novel, related knowledge (Markman & Hutchinson, 1984), and this learning benefit is associated with hippocampal and medial prefrontal processing (Tse et al., 2007, 2011; van Kesteren, Fernández, et al., 2010; van Kesteren et al., 2014). Without prior constraints, however, humans typically require many trials to acquire generalized knowledge structures (Seger et al., 2000; Seger & Cincotta, 2006). In contrast with the hippocampal memory system that stores event-specific information, the neocortex accumulates generalized knowledge across multiple similar events to build the general statistical structure of environmental relations (McClelland et al.,

1995; O'Reilly & Norman, 2002). A similar relationship is found in corticostriatal loops during feedback-based learning of stimulus-response associations, with the basal ganglia responding on a trial-by-trial basis to reward-prediction-errors (reward-gated plasticity) and the neocortex responding to the accumulation of reward across repetitions (reward-shaded plasticity; (Seger & Miller, 2010)). Both the striatum and hippocampus thus learn on the basis of unique exposures to exemplars (Daw, Niv, & Dayan, 2005; Seger & Cincotta, 2006) in contrast with a slower learning neocortex (Pasupathy & Miller, 2005). The distinct contribution of the striatum and the hippocampus is often subtle as demonstrated by research on the probabilistic classification task. Here, patient studies suggest learning to be dependent on the striatum (Dalton et al., 2013; Holl et al., 2012) and the hippocampus (Knowlton et al., 1994). Imaging studies report either a trade-off with initial hippocampal activity and a slower build-up of prolonged activity in the caudate nucleus of the striatum (Poldrack et al., 2001), or both striatal and hippocampal activity tracking initial learning (Kumaran et al., 2009). In the latter study, the pattern across many single associations contained a higher-order structure consisting of two associative rules, and generalization of this structure was related to initial learning-related activity in the hippocampus and connectivity between hippocampus and medial prefrontal cortex (Kumaran et al., 2009). Other imaging studies of initial knowledge acquisition have also typically used tasks with one or two associative rules and dichotomous choice options (Seger et al., 2000; Seger & Cincotta, 2006). However, from these studies it is not clear how activity dynamically unfolds across learning in the striatum, hippocampus and neocortical knowledge representation areas.

This fMRI study uses a novel learning paradigm consisting of a linguistic hierarchical knowledge structure that is gradually acquired (inspired by Kirby et al. 2008), allowing us to model trial-by-trial knowledge build-up across a number of interleaved trials within a single learning session. Critically, participants learned associations between geometric figures on the one hand (defined by their color, shape and movement) and tri-syllabic word strings on the other hand. These associations contained higher-order regularities in their mappings across exemplars. Exemplar associations were replaced with new exemplars halfway through the learning session, allowing the establishment of generalization

across trials (see figure 1). We then tracked learning from an initial state, where associations are deemed arbitrary, to an end-state where these associations have acquired meaning within the higher-order associative structure. The State-Space model (Smith et al., 2004) was then used to systematically track accumulation and updating of knowledge during the acquisition session during the cue presentation (cue phase) and feedback presentation (learning phase). This approach allows us to closely track how various brain regions dynamically contribute to the gradual acquisition of a complex linguistic knowledge structure.

## Materials and methods



### Participants

Thirty-two healthy, right-handed subjects with normal or corrected-to-normal vision participated in the experiment (age range: 19-32 years; 20 female). Subjects received monetary compensation for participation and could earn extra money based on performance. One subject was excluded from further analysis due to scanner malfunction, and five subjects were excluded because they failed to reach the learning criterion, defined as reaching the critical learning trial (CLT, see Behavioral Analysis) over the course of the experiment. One further participant was excluded from specifically the MRI-analysis due to excessive movement inside the scanner. All subjects provided written informed consent. The study was conducted according to a protocol approved by the local review board (CMO Region Arnhem-Nijmegen, the Netherlands).

### Stimuli

The stimuli consisted of geometric figures described by artificial tri-syllabic pseudo-word labels. The figures and pseudo-word labels did not have a meaningful association to real-life figures and words. As such, the influence of prior knowledge on learning the associations was minimized. More specifically, the stimuli consisted of visual geometric figures defined by a specific color, shape, and movement trajectory across the screen. There were eight isoluminant colors (red, green, blue, white, yellow, cyan, magenta, and black) presented on an isoluminant grey background, eight different shapes (square, circular, vertical rectangle, star, diamonds, vertical rectangle, hexagon, and triangle, all made to fit into a

rectangle of 142 pixels), and eight movement trajectories (all moving at two pixels per screen refresh along a polar angle of respectively 0, 45, 90, 135, 180, 225, 270 and 315 degrees). In total, these three features could be combined into 512 unique exemplars. Similarly, 512 unique pseudo-word labels could be made on the basis of the 24 syllables used (see Figure 4.1).

Two sets of associations were used for the acquisition session. Both sets contained four colors, shapes, and movements and each particular perceptual feature was only used in one of the two sets. Therefore, each set contained 64 exemplars with a unique configuration of three perceptual features (color, shape and movement). The syllables were distributed across both sets, each set containing 64 pseudo-word exemplars with a unique configuration of three syllables. The regular set contained associative regularities across exemplars that could be discovered and generalized to novel exemplars. Each visual feature was consistently associated with a syllable in a fixed position within the tri-syllabic pseudo-word (color corresponded to the first, shape to the second, and movement to the third syllable). Furthermore, each instantiation of a feature corresponded to an uppercase syllable (e.g. red = 'NI', square geometry = 'SA', rightward movement = 'ZA', Figure 4.1). In contrast, the irregular set contained no regularity across individual exemplars. Here, in different exemplars, features (colors, shapes and movements) were paired with different syllable locations, and feature instantiations (e.g. 'red', 'square' and 'rightward movement') were paired with different syllables. The pairings between figure and pseudo-word labels were consistent only within repetitions of the same exemplar, but not in different exemplars with overlapping features (see Figure 4.1 for examples). colors, geometric shapes, movements and syllables were assigned to the two sets in a counterbalanced manner across participants.

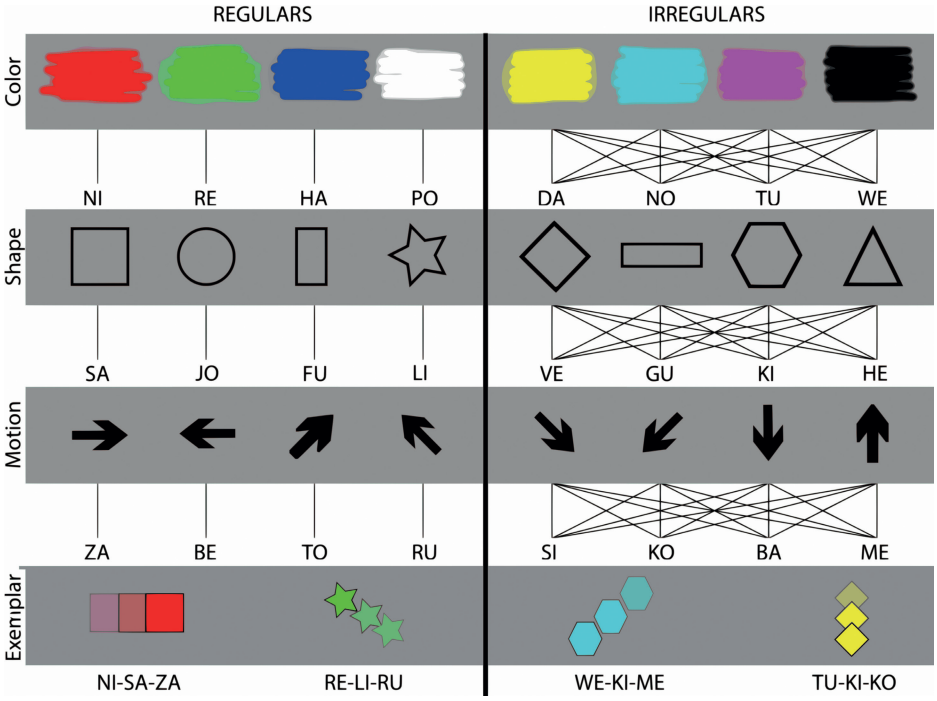



**Figure 4.1. Stimulus materials.**

*Subjects were presented with pairs of geometric figures and artificial pseudo-word labels. The figures consisted of a defining color, shape and movement, and the pseudo-words consisted of three corresponding syllables that uniquely mapped onto the figures' perceptual features. In the regular set, a particular instantiation of a perceptual feature was always associated with the same syllable (left panel), whereas in the irregular set the same feature instantiation could be associated with different syllables, depending on the exemplar configuration (right panel). Below: four examples of associations between figures and artificial pseudo-words from the regular set (left panel) and the irregular set (right panel) are shown.*

### Task and procedures

Participants were instructed that they were to learn a new language consisting of geometric figures denoted by a tri-syllabic pseudo-word label. They were informed that there was regularity in the associations between figures and artificial pseudo-words, but the nature of this regularity was not disclosed. Participants viewed blocks consisting of either geometric figures or pseudo-word labels, such that all perceptual features and syllables had been seen across this pre-experimental exposure period. This served to familiarize participants with the stimuli on a perceptual level and to rule out stimulus novelty effects during the ensuing scan session. To become familiarized with the requirements of the task, participants

were briefly trained on nine trials similar to those that were presented later during the acquisition session. This training used colors, shapes, movements and syllables that were randomly sampled from the entire set. During the scanned acquisition session participants were presented with trials from both regular and irregular sets distributed across 8 blocks (Figure 4.2). During the first four blocks, twelve regular figure-pseudo-word combinations and four irregular figure-pseudo-word combinations were repeated in each block. In the second set of four blocks, 12 new regular and four new irregular combinations were introduced and repeated across the remaining blocks. As such, participants' ability to generalize across exemplars of block 4 and 5 could be assessed. The order of regular and irregular trials within a block was randomized. In total, 96 trials from the regular set and 32 trials from the irregular set were presented across four runs (each run contained two blocks). In each trial, the figure was presented for 3s (the retrieval-phase), then participants were asked to select the corresponding tri-syllabic pseudo-word from three options per syllable (the response-phase, duration 6s), and lastly the correct response was presented (feedback-phase, duration 3s). Participants were instructed to retrieve the pseudo-word label immediately upon seeing the figure and to maintain that information until they could give the correct response (Figure 4.2).

During the response-phase, participants were asked to select three syllables comprising the complete word, by selecting one out of three alternatives for each syllable location (first, second, and third syllable of the word). The chance level for having one individual syllable correct is therefore 1/3, and the chance level for having the entire word correct is 1/27 (1/3 * 1/3 * 1/3). A blue bar underlined the syllable location where at each moment a syllable needed to be selected (first, second, and third syllable of the word). To reduce interference across sets, the irregular set was clearly distinguishable from the regular set. Specifically, the three syllable alternatives were underlined with light-blue bars, such that participants were able to distinguish the regular and irregular sets. The regular and irregular trials were similar, but participants could not generalize learned associations across exemplars in the irregular set. For the alternative response options, syllables were pseudo-randomly selected from all syllables that occurred across regular and irregular associations at that particular syllable location (first,

second, and third syllable of the word). This ensured that both syllables from the regular and irregular set were repeated to a roughly equal extent across learning, and effects of syllable familiarity were ruled out.

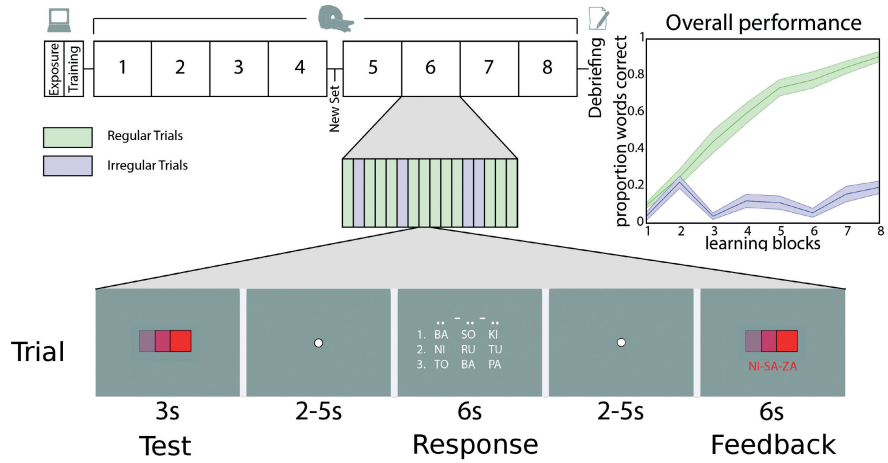



**Figure 4.2. Task design and task performance.**

*After pre-exposure to stimulus materials and pre-training, participants learned figure-pseudo-word associations across eight blocks. Each block contained 12 regular trials and 4 irregular trials, these were repeated four times. After the first four blocks, the exemplars were switched. In each trial, first the figure was presented (the test-phase, duration 3s), then the corresponding tri-syllabic pseudo-word was selected (the response-phase, duration 6s), and lastly the correct response was presented (feedback-phase, duration 3s). Participants were instructed to retrieve the pseudo-word label upon seeing the cue, and maintain that information until the correct response could be given. Inset: Plot of overall performance for irregular and regular trials demonstrates a steadily improving learning curve for regular trials.*

In the feedback-phase, the figure was presented again with the correct pseudo-word underneath, presented in red if one of the selected syllables had been incorrect and green if the entire pseudo-word was correct. Participants were instructed to compare feedback with their initial response and learn from the discrepancies between response and feedback. All trials and phases within a trial were separated by a jittered interval of 2-5 seconds. During the entire experiment the participants were instructed to fixate on a white fixation dot that remained visible at the centre of the screen. After each block, a baseline block with a jittered length of 9-11 seconds was presented. The stimuli were presented using a projector at the rear of the scanner bore (60 Hz refresh rate, 1024 by 768 resolution) viewed by the participants through a mirror attached to the headcoil (covering 6° of

horizontal and 7° of vertical visual angle). The experiment was programmed in Matlab, using the Psychophysics Toolbox extensions (Brainard, 1997).

Following completion of the acquisition session, participants were debriefed to assess whether they had discovered the organizing principles underlying the associations. They were first asked to provide the meaning of all 24 syllables from the regular set (e.g. by writing 'red' for color, writing 'hexagon' or making a small drawing for shape, and 'to the right' or an arrow for movement). Next, they were asked whether they knew which perceptual feature corresponded to the first, second, and third syllable in the pseudo-word label. As such, this questionnaire probed explicit knowledge of the associative regularity inherent in the regular set.

**Behavioral analysis**

In the current paradigm, each pseudo-word label consisted of a sequence of three syllables. The participant was presented with three response options per syllable (33 % chance probability of selecting a correct syllable) and thus with a total of 27 possible pseudo-word responses per trial (~4% chance probability of correct response for the entire pseudo-word). Vectors coding (0 to 3 for the number of features correct) for trial-by-trial performance across all 128 trials were extracted for individual subjects.

To track the state of knowledge across the acquisition session, individual learning curves were estimated using a State-Space model (Smith et al., 2004). In this model the accumulation of knowledge is described as an increase in the probability of a correct response across trials. The model is characterized by two equations: an observation equation and a state equation. The observation equation describes how the observed binary choice data (e.g. 'correct' or 'incorrect') relates to a hidden state or latent learning process. Here, each trial involves a sequence of three responses to select the correct syllable for each perceptual feature. Therefore, the observation equation is best characterized as a binomial process (i.e. a sequence of three Bernoulli trials). The state equation describes the hidden learning process that evolves across trials and is defined as a Gaussian random-walk process. Therefore, learning in this model is reflected by an increase in the latent state process that ultimately leads to an increase in the number of correct responses

(i.e. higher probability of selecting the correct syllable for each feature). In a state-space model, inferences on the learning process are made from the perspective of an ideal observer, using the complete sequence of trials to estimate the time-course of learning. A state-space model was fitted to the data using numerical optimization techniques based on the Expectation Maximization (EM) algorithm (code obtained from www.neurostat.mit.edu).

To verify that the state-space model provided the best account of the al data, three alternative models were also fitted to the learning data: the Rescorla-Wagner model (Rescorla & Wagner, 1972), the learning component of the memory chain model (MCM; Chessa & Murre, 2007) and the moving average model (MA; Smith et al., 2004). The models were fitted to the data using customized code adapted from freely available online resources and Maximum Likelihood fitting routines. Fitting of the state-space model was carried out using custom code for MATLAB® (adapted from Smith et al., 2004). Fitting of the Rescorla-Wagner model was done by calculating the equilibria of the model (Danks, 2003) using functions implemented in the R® package 'ndl' (Arppe et al., 2015). 2015). Fitting of the MCM model was carried out using custom code and standard optimization routines for Mathematica®. Last, the moving average model was fitted to the data using the toolbox Forecast (Hyndman et al., 2013) for R®.

Model selection was carried out by calculating Bayesian Information Criterion scores (Schwarz, 1978) for each model, which is a measure of the goodness-of-fit that penalizes for the number of free parameters. Therefore, the model with the lowest score represents the most parsimonious account of the data. In line with previous studies (Kumaran et al., 2009), results indicated that the state-space model provided the best description of the observed data (see Table 4.1). The fit of learning curves generated by the moving average method was worse than the state-space model, but superior to the Rescorla-Wagner model. The memory chain model performed better than the RW, but worse than the state-space model or moving-average model. The state-space model thus provides the most accurate description of the experimentally observed individual learning curves and therefore the estimated 'p mode' parameter (the mode of the distribution of estimated probabilities per trial) from the state-space-model was used as model

parameter for parametric model-based fMRI analyses. Furthermore, the critical learning trial (CLT, see Smith et al., 2004) was defined as the first trial where it can be concluded with reasonable certainty that a subject performs better than a certain threshold (95% confidence interval exceeds chance performance). This threshold performance was here defined as 66%, which corresponds roughly to having two syllables correct.

**Table 4.1.** Model comparison.

| Model | BIC | Description |
|---|---|---|
| RW | 545 | Rescorla-Wagner model |
| MCM | 252 | Memory Chain Model |
| MA | 234 | Moving Average model |
| STA | 62 | State-Space model |

*Bayesian Information Criterion (BIC) scores for each of the four learning models. The model with the lowest score represents the most parsimonious account of the data.*

### Imaging parameters and acquisition

T2*-weighted echo planar images (EPI) with BOLD (blood-oxygen-level-dependent) contrast were acquired on a 3 Tesla Siemens Skyra MRI scanner. We scanned 45 oblique axial slices angled at 30° in the anterior-posterior axis, TR 2.44 s, 2 mm thickness (0.5 mm gap), in-plane resolution 2.5 × 2.5 mm, field-of-view 212 mm. To minimize signal dropout in the temporal lobes and medial prefrontal cortex a dual-echo sequence (Halai, Parkes, & Welbourne, 2015) was used (TE1 = 15 ms and TE2 = 36). Additionally, a structural T1-weighted 3D magnetization prepared rapid acquisition (MPRAGE) gradient echo sequence image (192 slices, voxel size =1 × 1 × 1 mm) was acquired for each participant.

### fMRI data preprocessing

The first six volumes were discarded to permit T1 relaxation, and the next thirty volumes were used to calculate the weighting for recombining the two echo images. The images within a run were realigned to the first image, and subsequently the two echo images for each volume were recombined using custom scripts and weighting parameters (Poser, Versluis, Hoogduin, & Norris, 2006). Subsequently, images were preprocessed using the statistical parametric mapping software

SPM (http://www.fil.ion.ucl.ac.uk/spm/) and custom scripts written in Matlab (http://www.mathworks.com/products/matlab). The functional images were realigned across runs and coregistered to the structural image. Next, structural and functional images were segmented and normalized to MNI-space on the basis of their grey and white matter templates using DARTEL. Compartment maps were generated for grey matter, white matter and CSF, and non-specific nuisance regressors were created modelling signal from white matter and CSF-compartments. Normalized EPI images were smoothed using a Gaussian kernel with full-width half-maximum of 5 mm. Custom scripts were used to detect and remove spike artifacts.

**fMRI analyses**

The fMRI data was analyzed in SPM12 using the general linear model (GLM), estimated in two stages. Subject-specific experimental effects were modelled at the first-level using GLMs. Then, a random-effects analysis was performed at the second level using one-sample t-tests, resulting in group-level statistical parametric maps. Task regressors were included in GLMs to model the test-phase, response-phase and feedback-phase separately for each learning trial. The regular and irregular trials were also modeled separately. The inter-trial interval and baseline blocks served as implicit baseline. Events were modelled with a boxcar function with the same length as the respective phases (test-phase 3s, response- phase was modelled from presentation onset until the last keypress, or alternatively 6s if one or more responses was missing, feedback-phase 3s), and convolved with the canonical hemodynamic response function (HRF). Furthermore, button presses were modelled as stick functions and convolved with the canonical hemodynamic response function in a separate regressor. Movement parameters were included as regressors of no interest. To deal with non-specific signal fluctuations, signal time-courses were extracted from white-matter and cerebrospinal fluid compartments and included as regressors of no interest. We employed a high-pass filter with a low cut-off of 1/835 Hz, as the task power spectrum revealed that a more conventional cut-off of 1/128 Hz removed gradual learning-related fluctuations in the BOLD-signal. Runs were also concatenated to prevent the removal of gradual learning signals. Temporal autocorrelation was modelled using AR(1).

Analyses of regional activation focused on parametric modulations, namely regressors in GLMs to detect brain regions where BOLD-activity was modulated by the trial-by-trial state of knowledge during regular trials (the irregular trials were not considered here as the amount of trials was too low). The first model included subject-specific vectors denoting the probability of a correct response on any given trial (probability range: 0-1) as estimated by a state-space model (Smith et al., 2004). This vector is a proxy for the level of knowledge accumulated at any given trial, and thus approximates the amount of knowledge that could be retrieved at any trial (Figure 4.3A). The accumulation function was included as a parametric modulator in the GLM to model all phases of the regular trials. During the cue-period, a higher value approximates the amount of knowledge that can be retrieved. During the feedback-period, knowledge is not actively retrieved, and once more knowledge has accumulated less updating occurs. Therefore, activity related to accumulated knowledge during feedback presentation might represent a variety of factors (see Discussion). The second model included the difference function of the first regressor as parametric modulator (the state-space learning curve), which indicates the change in probability of a correct response on each trial compared to the previous trial (Figure 4.3B). During the cue phase, the updating parameter is an estimate for recently acquired knowledge that can be recruited in the current test-phase. This function was given the value of zero for the first trial, where no knowledge was present by definition. During the feedback-phase, the same updating function was used, but shifted by one trial earlier in the sequence (such that the updating function represents the change in knowledge on the current trial compared to the next), and the updating function was appended with a 0 in the last trial. Both model 1 and model 2 thus included 10 task regressors (test-phase, response-phase and feedback-phase for regular and irregular trials, button presses, parametric modulator for test-phase, response-phase and feedback-phase of regular trials), along with the compartment and movement regressors. Condition-specific experimental effects (regression coefficients) were obtained in a voxel-wise manner for each participant. At the second (random-effects) level, participant-specific linear contrasts of the parameter estimates were entered in a one-sample t-test to construct the group-level statistical map. We considered results that survived correction for multiple comparisons (family-wise error (FWE) cluster-level correction) at $p<0.05$ across the whole brain or within

independently defined anatomical masks (small-volume correction, SVC) after initial thresholding at p=0.001 uncorrected. Anatomical masks were defined using the WFU PickAtlas Tool 2.4 as implemented in SPM.

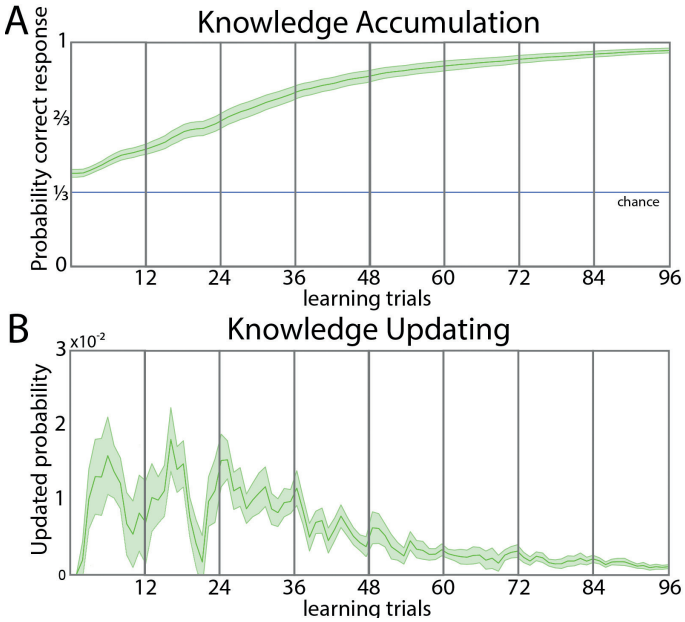



**Figure 4.3. Learning parameters estimated from the State-Space model on the regular trials.**

*Top panel: Group-averaged probability of making a correct response on each regular trial as estimated by the State-Space model. Bottom panel: Averaged change of the probability of making a correct response on each trial compared to the previous trial, as estimated by the State-Space model.*

### Beta-series extraction

To characterize the temporal dynamics of regional activity across learning, beta-values were analyzed from regions of interest across all learning trials. Specifically, trial-specific beta-images were obtained for all cue- and feedback-presentations by running one GLM with separate columns for each cue- and presentation. A second GLM was run with separate columns for each feedback-presentation. The resulting beta-images were sorted according to their temporal trial order for cue and feedback presentations, and signal was extracted across beta-images for several regions of interest, which were informed by the results from the whole-brain analyses. Furthermore, to control for subject-specific differences in the

learning rate, the timeseries were zero-centered around the critical learning trial. These subject-specific timeseries were then smoothed using a centered moving average with a sliding window of three timepoints, and the resulting vectors were averaged across all participants.

## Results

### Behavioral learning

Participants studied object-word associations across 128 trials in eight scanner blocks of 16 trials. Each block contained 16 unique associations that were repeated across the first four blocks, upon which the associations were changed and a different set of 16 associations was repeated across the next four blocks. Twenty-six participants reached the critical learning trial (mean CTL = 39.96, SD = 19.12), and exceeded chance level performance on the last block (proportion of all words correct on last block, mean = 0.74, SD = 0.11, chance level = 0.037, $t_{(25)} = 33.54$, $p < 0.001$, see Figure 4.2). These twenty-six participants were included in the final analyses. Each block contained twelve exemplars from the regular set and four exemplars from the irregular set. In both conditions, learning was evident by above-chance performance in the eighth and last block (proportion regular correct, mean = 0.92, SD = 0.12, versus chance level = 0.037, $t_{(25)} = 37.84$, $p < 0.001$, mean proportion irregular correct, mean = 0.19, SD = 0.18, versus chance level = 0.037, $t_{(25)} = 4.46$, $p < 0.001$). However, performance on regular trials was better than on irregular trials (mean proportion difference = 0.73, SD = 0.18, $t_{(25)} = 20.75$, $p < 0.001$). Consistency of mappings across exemplars in the regular set might benefit learning through generalization of mappings across exemplars, while reducing interference between similar exemplars. This was confirmed by comparing behavioral performance in the fourth to the fifth block, when the exemplars were changed. Indeed, performance on regular exemplars showed a consistent increase (proportion correct increase regulars, mean = 0.13, SD = 0.17, $t_{(25)} = 3.97$, $p < 0.001$), which is in line with generalization of acquired knowledge across different sets of exemplars. In contrast, there was no performance increase for irregular trials between block 4 and 5 (proportion correct increase irregulars, mean = 0.01, SD = 0.27, $t_{(25)} = 0.18$, $p = 0.86$) and the performance increase was larger for regular than irregular exemplars ($t_{(25)} = 2.11$, $p = 0.045$).

Further evidence of the learning of associative regularity is provided by analysing performance on specific features of the regular figures. By the eight block of acquisition on regular trials, all subjects managed to learn the syllables associated to color (mean colors correct: 11.85 out of 12, SD = 0.61, chance level = 4, $t_{(25)} = 65.30$, $p<0.001$), shape (mean correct: 11.69, SD = 0.84, $t_{(25)} = 46.83$, $p<0.001$), and motion (mean correct: 11.38, SD = 0.94, $t_{(25)} = 40.00$, $p<0.001$). When probing explicit knowledge of associative regularities at debriefing, all participants could indicate that color was associated with the first syllable, geometric shape with the second, and movement with the third syllable. Furthermore, when presented with syllables that had occurred in the regular set, along with lures, participants could indicate the meaning of the syllables from the regular set when explicitly probed (mean correct = 11.33 out of 12, SD = 1.52).



**Brain activity associated with feedback-based learning**

At the start of learning, participants had no access to prior knowledge of the associative knowledge structure. Therefore, they were required to use the information provided during feedback to update their knowledge. To probe brain areas associated with the updating of knowledge, individual learning parameters were fitted to functional MRI data. A State-Space model was used to estimate for each trial the amount of knowledge that was updated compared to the next trial (see Table 4.1 and Figure 4.3). Specifically, a large area (see Figure 4.4C and Table 4.2) spanning the left ventrolateral and lateral prefrontal cortex (MNI coordinates: -38 22 22, $p<0.001$ FWE corrected for the whole brain) was found to exhibit activity that covaried with the extent to which knowledge is updated based on feedback in the current trial. This region thus may contribute to the updating of representations of the associative knowledge structure. When modeling the accumulated knowledge during feedback, activity was found to covary in several large midline regions, including bilateral medial prefrontal cortex and posterior cingulate cortex, bilateral superior/middle temporal gyrus, and left-lateralized hippocampus, angular gyrus, and somatosensory cortex (see Figure 4.4D and Table 4.3). These regions were more active during feedback presentation when more knowledge had already been accumulated.

**Table 4.2.**

| Brain region | Cluster (voxels) | Laterality | Z-score | Local maxima | | |
|---|---|---|---|---|---|---|
| | | | | x | y | z |
| Lateral prefrontal cortex | 239 | L | 4.45 | -38 | 22 | 22 |

*Brain areas where activity significantly correlated trial-by-trial with recently updated knowledge during the feedback trials of the acquisition session. All coordinates are in MNI space.*

**Table 4.3.**

| Brain region | Cluster size (voxels) | Laterality | Z-score | Local maxima | | |
|---|---|---|---|---|---|---|
| | | | | x | y | z |
| Medial Prefrontal Cortex | 1740 | L/R | 6.45 | -8 | 55 | 2 |
| Inferior Parietal Cortex | 230 | L | 5.45 | -52 | -62 | 38 |
| Somatosensory Cortex | 260 | R | 5.34 | -65 | -22 | 38 |
| Inferior Parietal Cortex | | | 4.89 | 62 | -30 | 45 |
| Inferior Parietal Cortex | | | 4.32 | 55 | -32 | 40 |
| Posterior Cingulate Cortex | 853 | L/R | 5.26 | -2 | -12 | 38 |
| Middle Temporal Gyrus | 203 | L | 5.18 | -62 | -18 | -15 |
| Hippocampus | 85 | L | 4.75 | -25 | -18 | -18 |
| Middle/Superior Temporal Gyrus | 91 | R | 4.66 | 60 | -10 | -10 |

*Brain areas where activity significantly correlated trial-by-trial with accumulated knowledge during the feedback trials of the acquisition session. All coordinates are in MNI space.*

### Brain activity associated with cued retrieval

When presented with the figure during the test-phase, participants were asked to immediately recall the associated tri-syllabic pseudo-word. Participants updated their knowledge of the associations based on feedback, which they could then apply in the next test-phase. To probe which brain regions aid the deployment of this recently updated knowledge, the updated knowledge parameter was found to track brain activity specifically in the right caudate nucleus (peak MNI coordinates: 8 18 5, $p<0.05$ corrected for a reduced search volume in an anatomical mask defined by the right caudate nucleus). Thus, the caudate nucleus appears to be involved in the deployment of recently updated knowledge during the test-phase (see Figure 4.4A

and Table 4.4). Complementary to this, the knowledge accumulation parameter could be used to probe which regions covary with the total amount *of* knowledge that had accumulated across previous learning trials and could thus be recruited in the current trial. Several regions were found, including low-level and associative visual regions, primary, premotor and supplementary motor areas, and frontal and parietal cortices (see Figure 4.4B and Table 4.5). Importantly, the left inferior frontal gyrus that we had found to track feedback-based knowledge updating was found to also track accumulated knowledge during the test-phase (peak MNI coordinates: -50 12 32, $p<0.001$ FWE corrected for the whole brain, and peak MNI coordinates: -50 30 22, $p<0.001$ FWE corrected for the whole brain). In sum, we found a set of regions spanning visual and motor cortex, as well as frontoparietal regions that were more active when more accumulated knowledge was available to be recruited during the test-phase.



**Table 4.4.**

| Brain region | Cluster size (voxels) | Laterality | Z-score | Local maxima | | |
|---|---|---|---|---|---|---|
| | | | | x | y | z |
| Caudate nucleus | 8 | R | 4.01 | 8 | 15 | 5 |

*Brain areas where activity significantly correlated trial-by-trial with updated knowledge during the retrieval trials of the acquisition session. All coordinates are in MNI space.*

**Table 4.5.**

| Brain region | Cluster size (voxels) | Laterality | Z-score | Local maxima | | |
|---|---|---|---|---|---|---|
| | | | | x | y | z |
| Premotor Cortex | 460 | L | 5.60 | -48 | -8 | 50 |
| Inferior frontal gyrus | | | 5.00 | -50 | 12 | 32 |
| Posterior Cingulate Cortex | 1176 | L/R | 5.46 | 8 | -50 | 5 |
| Occipital Cortex | | | 5.32 | 2 | -80 | -5 |
| Retrosplenial Cortex | | | 4.32 | -8 | -58 | 8 |
| Suppl. Motor Area | 473 | L/R | 5.43 | -2 | 10 | 58 |
| Pre-Supp. Motor Area | | | 5.14 | 2 | 18 | 45 |
| Anterior Cingulate Cortex | | | 4.21 | -8 | 12 | 38 |
| Superior Parietal Cortex | 238 | L | 5.24 | -20 | -62 | 52 |
| Posterior Parietal Cortex | | | 4.29 | -18 | -82 | 40 |
| Intraparietal Sulcus | | | 3.70 | -28 | -65 | 35 |
| Fusiform Gyrus | 257 | L | 4.89 | -45 | -40 | -20 |
| Occipitotemporal Cortex | | | 4.76 | -48 | -60 | -12 |
| Fusiform Gyrus | | | 4.54 | -38 | -35 | -25 |
| Occipital Cortex | 73 | L | 4.69 | -35 | -85 | 18 |
| Supplementary Motor | | L | | -58 | 10 | 0 |
| Cortex / Broca's Area | 48 | | 4.44 | | | |
| Thalamus | 44 | L | 4.38 | -10 | -12 | 10 |
| Intraparietal Sulcus | | R | 5.35 | 18 | -65 | 55 |
| Occipitoparietal Cortex | | | 3.89 | 18 | -78 | 45 |
| Precuneus | 105 | | 3.74 | 20 | -70 | 38 |
| Somatosensory Cortex | | L | 4.27 | -58 | -18 | 20 |
| Auditory Cortex | 80 | | 3.57 | -62 | -22 | 10 |
| Inferior Parietal Cortex | | L | 3.96 | -55 | -30 | |
| Somatosensory Cortex | | | 3.91 | -50 | -30 | |
| Intraparietal Sulcus | 52 | | 3.68 | -45 | -4 | |

*Brain areas where activity significantly correlated trial-by-trial with accumulated knowledge during the retrieval trials of the acquisition session. All coordinates are in MNI space*

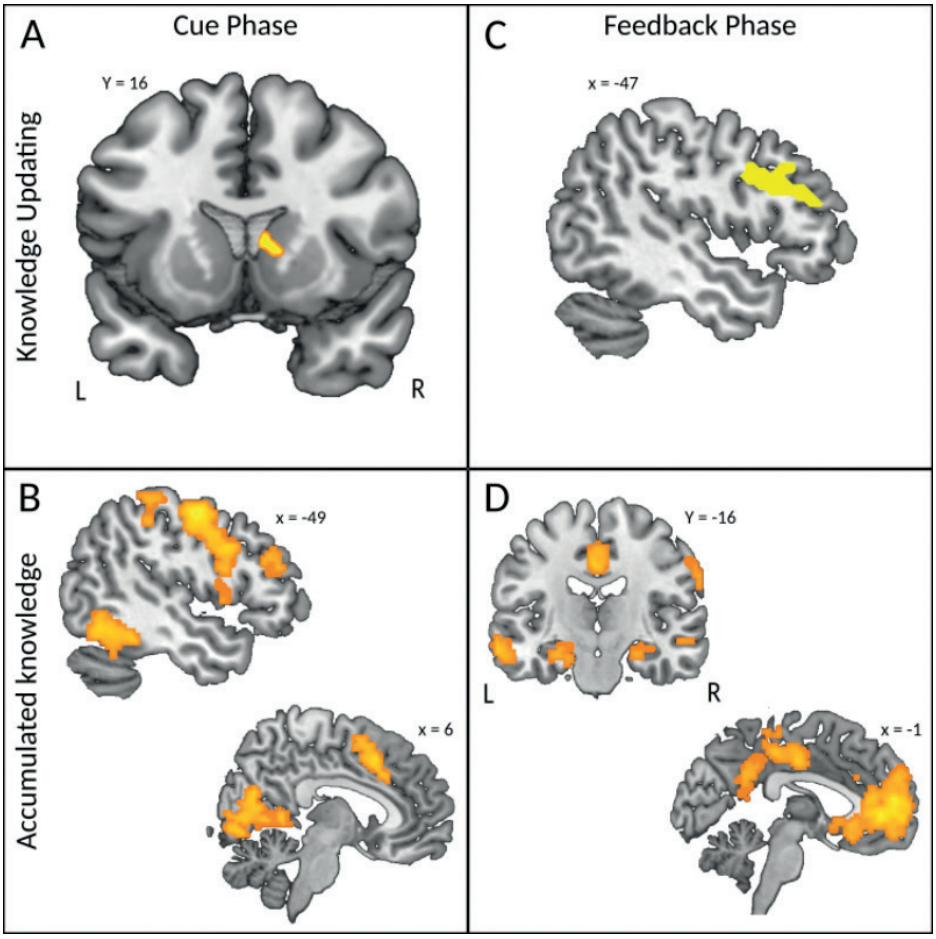



**Figure 4.4. Model-based parametric modulation analysis of learning-related activity**

*A) Top left panel. Activity in the right caudate nucleus shows a correlation with recently updated knowledge available at each test-phase.*

*B) Lower left panel. Regions where activity is correlated with accumulated knowledge at each test-phase, including left inferior frontal gyrus, and visual, motor, and frontoparietal regions.*

*C) Top right panel. Activity in the left ventral and lateral prefrontal cortex is correlated with updated knowledge during the feedback-phase.*

*D) Lower right panel. Regions where activity is correlated with the accumulated knowledge during the feedback-phase, including medial prefrontal cortex, posterior cingulate cortex, left hippocampus, and inferior parietal cortex.*

**Beta-series extraction**

We found regions that responded specifically to the knowledge updating parameters, namely the left inferior frontal gyrus and right caudate nucleus. These regions might thus be involved in the active updating and retrieval of knowledge. Furthermore, our model detected a wide variety of regions covarying with accumulated knowledge during feedback presentation. Of these regions, the left hippocampus is implicated in enabling the generalization across individual learning instances in interaction with the medial prefrontal cortex (Kumaran et al., 2009). It could be that these regions, which covaried with accumulated knowledge during feedback, had some role in generalization of knowledge during feedback-based learning, as more accumulated knowledge might benefit knowledge generalization. To visualize the relative unfolding of activity of these regions across both the test-phase and the feedback-phase, we extracted the beta time-series from these regions of interest (see Figure 4.5). The left hippocampus and right caudate nucleus were defined by anatomical masks, whereas functional regions found in respective contrast maps (thresholded at $p<0.001$ uncorrected) were used to define the left inferior frontal gyrus (based on feedback-based knowledge updating) and medial prefrontal cortex (feedback-based accumulated knowledge). The activity across trials estimated by this analysis is displayed in Figure 4.5, and some patterns become apparent upon visual inspection. During feedback-based learning, the left inferior frontal gyrus displays initially high activity, which levels off as more knowledge accumulates and less updating takes place. During cue-based retrieval, on the other hand, activity is initially low, but increases as more knowledge accumulates and becomes available to be retrieved. The caudate nucleus displays both initially high activity during feedback-based learning and cue-based retrieval, suggesting it might aid the initial updating and subsequent retrieval of knowledge. Both the hippocampus and medial prefrontal cortex display an increase in activity during feedback-based learning, corroborating the parametric modulation results. During cue-based retrieval, both the hippocampus and the medial prefrontal cortex display an initially high but gradually decreasing activity across retrieval trials as more knowledge is accumulated, suggesting these regions, contrary to expectations, do not become more involved with retrieval as more knowledgee accumulates.

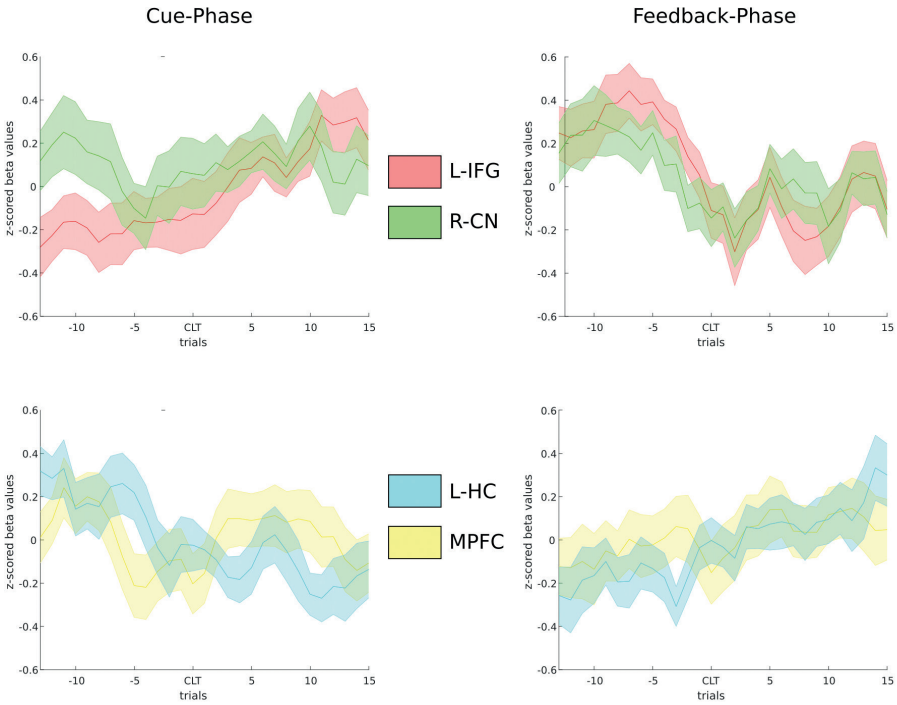



**Figure 4.5. Beta-series analysis of learning.**

*Beta-values were extracted during successive test (left column) or successive feedback presentations (right column) across learning, and plotted relative to the critical learning trial (CLT). Signal was extracted for the left inferior frontal gyrus (functional region that displayed modulation during feedback by knowledge updating; red colors), right caudate nucleus (anatomical mask, this region displayed modulation during the test-phase by recently updated knowledge; green colors) in top row. Signal was also extracted for the left hippocampus (anatomical mask, region displaying modulation during feedback-phase by accumulated knowledge, yellow colors) and medial prefrontal cortex (functional region that displayed modulation during feedback-phase by accumulated knowledge, blue colors. CLT = critical learning trial, L-IFG = left inferior frontal gyrus, R-CN = right caudate nucleus, L-HC = left hippocampus, MPFC = medial prefrontal cortex.*

## Discussion

Few studies so far have tracked how brain regions dynamically contribute to the gradual acquisition of a complex knowledge structure (Kumaran et al., 2009), and even fewer have disentangled the effects of retrieval and feedback during such learning tasks. The present study addressed this gap by systematically tracking brain activity while learners acquired a novel linguistic associative knowledge structure. The key findings of this study are that feedback-based knowledge updating is associated with activity in the left inferior frontal gyrus. Mirroring this result, the left inferior frontal gyrus displayed a gradual increase in activity during retrieval as more knowledge accumulated. A complementary signal was found in the right caudate nucleus, where activity during retrieval was found to correlate with the amount of recently updated knowledge that could be retrieved. The results are in line with a model in which the acquisition of linguistic associative knowledge is subserved by the left inferior frontal gyrus, which is initially being supported by fast-learning subcortical regions such as the caudate nucleus (Seger & Miller, 2010). The hippocampus and medial prefrontal cortex, two other regions generally implicated in acquiring generalized knowledge structures (Constantinescu, O'Reilly, & Behrens, 2016; Jai, Liu, Loback, Grossrubatscher, & Frank, 2018; Kumaran et al., 2009; Morrissey, Insel, & Takehara-Nishiuchi, 2017), covaried with accumulated knowledge during feedback presentation. This potentially suggests that as more knowledge accumulates, more generalized features are extracted through inferential computations performed by the medial prefrontal cortex and hippocampus.

The left inferior frontal gyrus is thus important for linguistic knowledge acquisition. A search on Neurosynth (http://neurosynth.org/) revealed that the peak of the region (x=-38, y=22, y=22) displays the strongest meta-analytical association with the terms "BA44" (z=5.58), "syntax" (z=5.52), "word" (z=4.93) and "semantic" (z=4.82). Indeed, this region has widely been suggested to be involved in language processing (Bookheimer, 2002; Friederici, 2002; Hagoort, 2003; Hagoort, Hald, Bastiaansen, & Petersson, 2004). It is also implicated in the learning of sequences (Forkstam, Hagoort, Fernández, Ingvar, & Petersson, 2006; Peigneux et al., 1999), and sequences that are hierarchically organized (Gelfand & Bookheimer, 2003; Petersson, Forkstam, & Ingvar, 2004). It has also been

argued that the area described here, overlapping with Broca's area, is involved in retrieving word information from memory and combining them in larger sentence units (Hagoort, 2005). The current experiment includes elements from all these prior studies, requiring participants to parse hierarchically organized syntactical sequences of feature-syllable associations based on repeated exposures to larger word units (during feedback-based learning), and retrieve word information from memory (during cue-based retrieval).

An interesting finding during retrieval is the initial striatal learning signal, where the right caudate nucleus covaried with the retrieval of recently updated knowledge. This phasic signal during retrieval was complemented by activity in neocortical regions that increased during retrieval as a function of the amount of knowledge that had accumulated. Most of the regions found here were visual, motor, or attentional regions, suggesting they were involved in preparing for the ensuing visuomotor responses and might have been unspecific to the retrieval of knowledge. However, interestingly, the left inferior frontal gyrus, the region that was found to update knowledge on the basis of feedback, was also one of the regions found to display an increase in activity when more accumulated knowledge was available for retrieval, suggesting this region has a function in both the active updating of knowledge and subsequent retrieval of accumulated knowledge. This is congruent with a model where this region actively stores the syntactic associative structure: the more this knowledge structure is updated based on feedback, the higher the activity would be in regions that store the knowledge structure, and the more knowledge about the linguistic structure has accumulated across previous trials when recruiting it during retrieval, the more this region would be activated too.

An explanation for a striatal rather than a hippocampal involvement during initial retrieval might be given by the nature of the learning task involved here. The caudate nucleus has generally been found to be important for reward-based reinforcement learning (Haruno et al., 2004; Haruno & Kawato, 2006). In a word learning task, signals in the ventral striatum have been related to an implicit or self-generated reward when the meaning of a novel word was acquired (Ripollés et al., 2015), while also showing strong connectivity with the more dorsal caudate

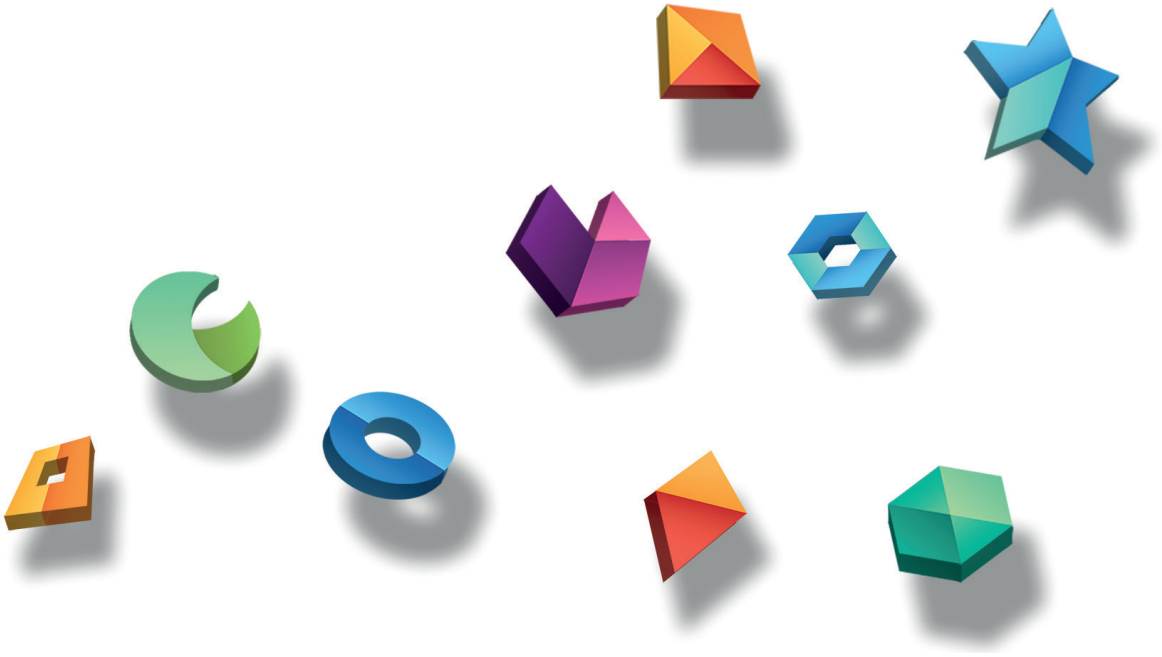
nucleus and the left inferior frontal gyrus during word learning. In our study, the updating signal observed during the retrieval of knowledge in the caudate nucleus may not have been an explicit reward signal itself (although the feedback provided was probably also experienced as rewarding). Rather, the caudate nucleus is more broadly involved in the excitation of correct action schemas and the selection of appropriate sub-goals based on an evaluation of action-outcome contingencies (Grahn, Parkinson, & Owen, 2008). Moreover, prior studies investigating the learning of spatial locations have shown that the hippocampus is involved in incidental configural learning, whereas the dorsal striatum is primarily involved in associative reinforcement learning of stimulus-response contingencies (Doeller & Burgess, 2008; Doeller, King, & Burgess, 2008; Lee, Duman, & Pittenger, 2008; McDonald & White, 1993). In our task, it might be that participants initially processed associations entirely based on configurations, but then learned to associate certain perceptual features of the figure (color, shape or movement) with a certain syllable at a particular location in the pseudo-word. This possibility hints at an intriguing trade-off between hippocampal-dependent and striatal-dependent learning, in line with earlier learning studies (albeit using simpler knowledge structures) showing an initial peak in activity in the medial temporal lobe and a somewhat slower build-up of activity in the caudate nucleus (Poldrack et al., 2001). It could be hypothesized that the initial process of discovering regularities requires a recurrent similarity computation by the hippocampus (Kumaran & McClelland, 2012). The uncovered component parts of the associative structure are subsequently established in stimulus-response associations between single features and syllables, supported by the striatum. The acquired linguistic associative knowledge structure could then ultimately be stored and retrieved from the left inferior frontal gyrus. Indeed, beta-time series extraction demonstrated an initial increase in retrieval-related activity in both the right caudate nucleus and the hippocampus, as well as the left inferior frontal gyrus, the latter of which continued to increase in activity across retrieval trials (see Figure 4.5).

The hippocampus and medial prefrontal cortex are thought to contribute to the acquisition of generalized knowledge (Constantinescu et al., 2016; Jai et al., 2018; Kumaran et al., 2009; Morrissey et al., 2017). Kumaran et al. (2009) found that

activity in these regions covaries with accumulated knowledge during learning, and that this learning-related performance was predictive of performance on a transfer test. However, they did not distinguish between retrieval of knowledge when presented with a cue, and the subsequent updating of knowledge based on the provided feedback. In our report, we find that the accumulated knowledge covaries with activity in the hippocampus, medial prefrontal cortex and posterior cingulate cortex, similar to the report by Kumaran et al. However, we find this relation specifically during the presentation of feedback. Even though the associative structure acquired in our task consisted of linguistic material and is more complex, the basic computations needed to acquire the associative knowledge structure appear similar (Kumaran & McClelland, 2012). Thus, despite the fact that recurrent similarity computations subserved by the hippocampus and its connectivity with the medial prefrontal cortex might already initially take place to enable an acceleration of learning, these computations are continuously needed to acquire all associative rules (associations between single feature and syllables, and the larger hierarchical associative structure) until an optimal performance is reached. Thus, activity in the medial prefrontal cortex and the hippocampus would track recurrent similarity computations underlying generalization of knowledge, which would track accumulated knowledge during feedback presentation. Alternative factors might be postulated to explain a relation of these brain regions and accumulated knowledge, such as reward processing: the accumulated knowledge parameter scales with the correctness of feedback, potentially thereby eliciting an implicit reward signal. It could also be that participants became more disengaged from the task during feedback presentation as more knowledge had accumulated, as participants did not need to update their knowledge anymore based on the feedback. Moreover, various other regions in posterior cingulate cortex, temporal and parietal cortices were also found to covary with accumulated knowledge. Indeed, the regions found overlap to a significant extent with regions involved in reward processing, particularly the medial prefrontal cortex (McClure, Laibson, Loewenstein, & Cohen, 2004; Tom, Fox, Trepel, & Poldrack, 2007), and default mode networks active during task disengagement (Greicius et al., 2003; Raichle et al., 2001). More detailed paradigms are needed that can be performed in single scanned learning sessions to better assess the neural dynamics of these learning systems.

A cautionary note should be made. It could be argued that the learning benefit for regular trials is not solely due to an ability to generalize knowledge across trials. For instance, in irregular trials, feature-syllable associations are constantly changing, potentially creating interference between associations that partially share the same features. However, increased interference across trials is an inherent outcome of reduced consistency. Similarly, consistency across trials reduces interference, and therefore inherently promotes generalization across trials. Furthermore, we did not primarily establish that generalization took place by comparing learning performance in regular and irregular trials, we also find that participants generalized between the first half and the second half of the learning session when exemplars were switched.

In conclusion, this study provided novel insights into how linguistic associative knowledge is acquired by systematically tracking schematic knowledge formation while participants were learning an abstract artificial language organized by higher-order associative regularity. During learning, we found signals in the left inferior frontal gyrus, which responded both to feedback-based knowledge updating, and available accumulated knowledge during retrieval. A complementary signal was found in the caudate nucleus, where activity correlated with the availability of recently acquired knowledge during retrieval, suggesting it initially supports the retrieval of knowledge. Furthermore, we found that activity in a set of regions, including the medial prefrontal cortex and hippocampus, scaled with accumulated knowledge during feedback presentation, which might be indicative of increased generalization of features of the hierarchical knowledge structure. Together, these results provide a mechanistic insight into how linguistic associative knowledge is acquired by generalization across repeated learning experiences.



# CHAPTER 5

## Cued reactivation during slow-wave sleep induces connectivity changes related to memory stabilization

This chapter is based on:

Berkers, R.M.W.J., Ekman, M., van Dongen, E.V., Takashima, A., Barth, M., Paller, K.A., & Fernández, G. (2018). Cued reactivation during slow-wave sleep induces connectivity changes related to memory stabilization. *Scientific reports*, *8*(1), 16958.

## Abstract

Memory reprocessing following acquisition enhances memory consolidation. Specifically, neural activity during encoding is thought to be 'replayed' during subsequent slow-wave sleep. Such memory replay is thought to contribute to the functional reorganization of neural memory traces. In particular, memory replay may facilitate the exchange of information across brain regions by inducing a reconfiguration of connectivity across the brain. Memory reactivation can be induced by external cues through a procedure known as "targeted memory reactivation". Here, we analyzed data from a published study where auditory cues were used to reactivate visual object-location memories during slow-wave sleep. We characterized effects of memory reactivation on brain network connectivity using graph-theory. We found that cue presentation during slow-wave sleep increased global network integration of occipital cortex, a visual region that was also active during the retrieval of object locations. Although cueing did not have an overall beneficial effect on the retention of cued versus uncued associations, individual differences in overnight memory stabilization were related to enhanced network integration of the occipital cortex. Furthermore, the occipital cortex displayed enhanced connectivity with mnemonic regions, namely the hippocampus, parahippocampal gyrus, thalamus and medial prefrontal cortex, during cue sound presentation. Together, these results suggest a neural mechanism where cue-induced replay during sleep increases the integration of task-relevant perceptual regions with mnemonic regions. This cross-regional integration may be instrumental for the consolidation and long-term storage of enduring memories.

## Introduction

The neural representations of recently acquired declarative memories are thought to be selectively reactivated during ensuing rest periods, particularly during slow-wave sleep (Diekelmann & Born, 2010; Diekelmann et al., 2011). For instance, hippocampal place cells in rats that co-activated during active were found to be reactivated in concert during ensuing slow-wave sleep (Wilson & McNaughton, 1994). Later studies found that this coordinated reactivation, also dubbed replay, extended to the perceptual cortex involved in the initial processing of the stimulus, such as the visual cortex (Ji & Wilson, 2007). Furthermore, evidence of neural replay has also been found in the medial prefrontal cortex (Euston et al., 2007; Genzel & Battaglia, 2017; Peyrache et al., 2009). These results demonstrate the occurrence of neural memory replay in rodents, which has also been shown to be behaviorally relevant for the retention of associative memories (Dupret, O'Neill, Pleydell-Bouverie, & Csicsvari, 2010).



Neural memory reactivation is difficult to observe in humans, as its timing is generally unknown. However, reactivation can also be induced using cues associated with information learned previously, using "targeted memory reactivation". For example, presenting an odor during slow-wave sleep induces activation of the hippocampus and stabilizes associated memory traces (Rasch et al., 2007). Furthermore, whereas studies have found that audio-visual-spatial associations can be stabilized in a targeted manner using specific auditory cues (Creery, Oudiette, Antony, & Paller, 2015; Rudoy et al., 2009), these cueing effects can also affect all associations acquired within the same learning context (Oudiette, Antony, Creery, & Paller, 2013). Van Dongen and colleagues (2012) performed targeted memory reactivation in the magnetic resonance (MR) scanner using auditory cues that were paired with specific object-location associations. Even though no consistent effect of cueing on memory stabilization was found, individual differences in the effect of cueing positively correlated with activity in the hippocampus, thalamus and cerebellum during slow-wave sleep. Given repeated earlier findings of memory improvement due to auditory cues presented during slow-wave sleep (Ong et al., 2016; Oudiette et al., 2013; Schouten, Pereira, Tops, & Louzada, 2017), it is likely that the noisy scanner environment contributed to excessive variability here, with some individuals effectively tuning out all

auditory input, or alternatively that cueing affected all associations acquired within the same learning context (van Dongen, Takashima, et al., 2012).

Here, we re-analyze data acquired by van Dongen and colleagues (2012) using a principled graph theoretical analysis approach to characterize whole-brain connectivity changes. Triggering reactivation of memory traces with auditory cues could induce a process akin to spontaneous memory replay, which likely involves whole-brain connectivity changes facilitated by synchronous brain states during slow wave sleep (Bergmann, Mölle, Diedrichs, Born, & Siebner, 2012; Staresina et al., 2015). In this study, we assess whether those effects are reflected in shifts in functional connectivity patterns in response to exogenous (cued) memory reactivation, and whether these effects may be related to systems consolidation. Conventional analysis methods are often unable to capture the coordinated whole-brain connectivity changes that are expected during memory replay. Thus, a network perspective can be leveraged using methods developed within the framework of graph theory (Bullmore & Sporns, 2009). The *participation coefficient* is a metric that captures the extent to which a region integrates and distributes information as a result of the number and positioning of their contacts in the network (Power et al., 2013). Specifically, it quantifies network integration by looking at the importance of a given node for interactions between subnetworks, by measuring the connectivity of a region within a module compared to the connectivity to other modules. The participation coefficient can, as such, be used to index the integration of a given voxel (node) in the wider brain (network) during cued reactivation of visuospatial associations during slow-wave sleep.

If a cue induces memory reactivation akin to replay, and this replay indeed plays a role in memory consolidation involving a hippocampal-neocortical trace shift, then one should expect a coordinated neural activation of multiple regions, including the sensory features of the cued memory (in this case visual object features in the occipital cortex (DeYoe et al., 1996) and the spatial layout in the parahippocampal gyrus (Aminoff, Gronau, & Bar, 2007) as well as key mnemonic regions (such as the hippocampus, thalamus, and medial prefrontal cortex). The increased integration of these and other regions would be expected to be related to memory stabilization and an increased involvement of the neocortex in the memory trace across sleep.

## Materials and methods

The Materials and Methods have been described extensively in the original report (van Dongen, Takashima, et al., 2012). Here, we summarize relevant details about the experimental setup, supplemented with details about the data analysis performed for this report.

### Participants

This re-analysis includes data from all 22 participants included in the original report (see for complete description). Briefly, 56 participants were initially recruited from the online research participant system of the Radboud University Nijmegen (age range 18–27; 14 males). Twenty-two participants in the sample reached a sufficient amount of the slow-wave sleep stage in the scanner to allow at least 80 % of the cueing protocol to be completed whilst in this sleep stage. Furthermore, these participants did not display micro-arousals in response to sound presentations, nor did they report explicit knowledge of the sounds having been presented while they slept. The experiment was conducted in accordance with national legislation for the protection of human volunteers in nonclinical research settings and the Helsinki Declaration and approved by the local ethics committee (CMO Arnhem-Nijmegen, Radboud University Medical Center). Participants provided written informed consent prior to participation and were compensated for participation with course credits or a monetary fee.

### Procedures

The experiment started between 7 and 8 PM, and participants went to sleep between 10 PM and midnight (see Figure 5.1). Participants were placed in the MR scanner, and the sound volume was individually calibrated to a level at which the participant could distinguish individual sounds while the scanner was operating. Participants learned 50 object-location associations, each object-location pair being presented together with a unique sound, inside the scanner. When all associations were learned to criterion (all objects placed within 4 cm from the correct location on two subsequent rounds), a pre-test was performed on all associations (Test 1). Next, participants were prepared outside the scanner for polysomnographic recordings, and placed in the scanner again for a 2 hr rest/sleep period, while electroencephalograph (EEG) was monitored online and fMRI

data was continuously acquired. When participants had entered a stable slow-wave state for a period of time as visible on the ongoing polysomnography (Iber et al., 2007), sounds were presented to the participants using an MR-compatible headphone. Specifically, half of the sounds that were associated with specific object-location associations during learning were also presented as 'cue sounds' during sleep (25 sounds presented twice, for a total of 50 cue sound presentations). Moreover, control sounds that had not been previously associated with learning materials were presented (5 sounds presented five times, for a total of 25 control sound presentations). Cue and control sounds were grouped in blocks of five 500 ms presentations, each separated by a 4-8 s jittered inter-stimulus interval (total block duration: 32 seconds). The total amount of trial presentations was limited due to inherent practical difficulties with obtaining sufficient SWS-periods to play the sound recordings. Therefore, with this limitation in mind, a 2:1 ratio of cue trials versus control trials was implemented to maximize the likelihood to find an effect of reactivation on memory stabilization. After a 2 hr sleep, participants were awoken and subsequently were tested on all object-location associations (Test 2).

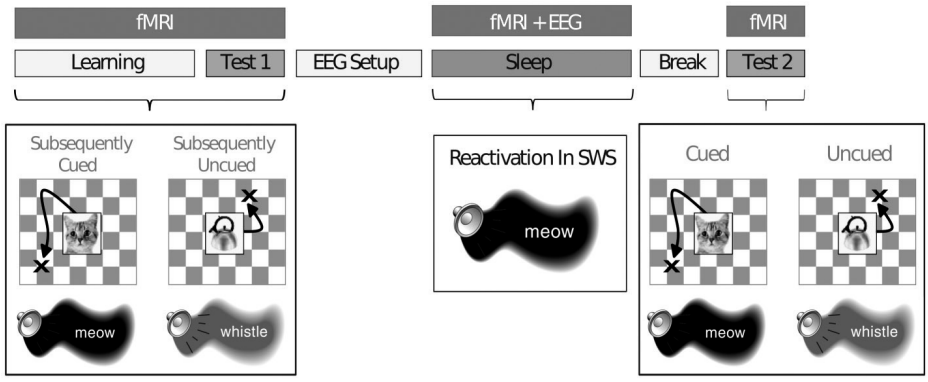



**Figure 5.1. Schematic depiction of the experimental procedure.**
*Participants learned 50 object-location associations, presented simultaneously along with object-related sounds, inside the MR-scanner. After a baseline test (Test 1), participants were set up for polysomnographic recordings and went to sleep inside the MR-scanner. Half of the learned associations were cued with sounds during slow-wave sleep. After awaking and a short break, participants performed another post-sleep test (Test 2). Figure adapted with permission from the initial report (van Dongen, Takashima, et al., 2012).*

**Object-location test**

The object-location task was performed in three stages: learning, Test 1 and Test 2. Here, we focus on the behavioral data from Test 1 and Test 2 for further analysis. Participants learned the location of 50 object pictures. For each participant, 50 objects were randomly assigned to 50 screen locations (screen size: 47 × 35 cm, resolution: 1024 × 768, viewing distance: 60 cm). In the learning phase, participants first passively viewed all 50 objects in their respective screen locations (duration: 3s presentation followed by a 1s inter-stimulus interval), paired with hearing the object related sound (e.g. a cat's 'meow' when the object was a cat, sound duration: 500 ms). Next, the learning phase continued with several iterative rounds of active learning, where an object was presented at the center of the screen simultaneously with the auditory stimulus. Participants were then required to place the object in its original location using an MR-compatible joystick, and to press a button to confirm the object placement (self-paced timing). Feedback was given for each trial by displaying the object in the correct location for 3s after a response was made. Objects were presented in a random order, and with every new round those objects that had been placed within 4 cm from the correct location on the two preceding rounds were excluded. Upon reaching criterion performance on all objects, participants performed the pre-sleep test (Test 1) for all objects, but without feedback. An identical post-sleep test was performed after sleep (Test 2). During Test 2, half of the object-location associations had previously been cued during sleep ('cued associations'), whereas the other half of the associations had not been previously cued ('uncued associations'). During tests, object placement was followed by an inter-stimulus period of fixation (duration jittered between 3 and 5s).

**MRI data acquisition**

Participants were scanned using a reduced-noise Echo-Planar Imaging (EPI) sequence with sinusoidal gradients to avoid acoustic resonances of the scanner (Schmitter et al., 2008). Functional (T2*) images were acquired with whole-brain coverage (28 axial slices, ascending slice acquisition, repetition time (TR) = 2,511 ms, echo time (TE) = 38 ms, 90° flip angle, matrix = 64 x 64, bandwidth = 1,502 Hz per voxel, slice thickness = 3.5 mm, slice gap = 15%, field of view (FOV) = 244 mm). Structural (T1) images were acquired with a magnetization-prepared rapid

acquisition gradient echo sequence (176 sagittal slices, TR = 2,250 ms, TE = 2.95 ms, 15° flip angle, matrix = 256 x 256, slice thickness = 1.0 mm, FOV = 256 mm).

**MRI data preprocessing**

The fMRI data acquired during Test 1, the sleeping period and Test 2 were preprocessed using standard routines implemented in SPM8 (www.fil.ion.ucl.ac.uk). The first five volumes of each functional EPI run were discarded, and an outlier algorithm was used to check for corrupted slices from each image that were replaced using between-volume interpolation. The functional images were then realigned, and coregistered to the structural image, spatially normalized to the Montreal Neurological Institute (MNI) EPI template (resampled at voxel size 2 x 2 x 2 mm), and smoothed using a Gaussian kernel (8 mm full-width at half maximum, FWHM). The structural images were segmented into gray matter, white matter, cerebrospinal fluid, and residual compartments (outside brain and skull) using the unified segmentation algorithm as implemented in SPM8. These compartments were used as masks to extract the mean intensity level across the whole time-series, and entered as compartment regressors to account for effects related to non-specific signal fluctuations.

**Network analysis during cueing**

We employed a graph theoretic framework (Bullmore & Sporns, 2009; Ekman et al., 2012) to analyze connectivity patterns during the presentation of sound cues in stable periods of slow-wave sleep. The aim was to measure dynamic changes in the whole-brain network properties during memory reactivation (by contrasting the presentation of cue sounds versus control sounds), and specifically changes in integration of certain brain regions with the whole brain network. Task-related connectivity was estimated using the common beta time-series approach (Rissman et al., 2004). Beta time-series are obtained by convolving each trial with a HRF separately (a separate regressor is included for each trial) and fitting the BOLD time-series using a standard GLM using FSL's Feat (including voxel-wise pre-whitening and AR(1) estimation). Resulting trial-specific beta estimates are then concatenated over trials to form a beta time-series. These beta time-series as such represent task-specific fluctuations.

A high-pass filter with a cutoff period of 128 s was applied to the functional time series. Sound presentation during sleep was modeled using separate regressors for each sound presentation (to obtain trial-specific beta-estimates for the 50 cue sound presentations and 25 control sound presentations, respectively) with a duration of 5 seconds each. We used a duration of 5 seconds to model cue presentation (which lasted 500 ms itself) to account for timing uncertainties with regards to the onset, duration and lag of the induced reactivation of the memory trace in response to the cue. First, the induction of memory reactivation by auditory input might depend on specific spindle-ripple events occurring during cortical up-states (Bergmann et al., 2012; Marshall & Born, 2007; Staresina et al., 2015). It is therefore difficult to be very precise with regards to a specific time-point where the auditory signal could potentially induce memory reactivation. Furthermore, once the auditory stimulus has reached the cortex and has induced reactivation of the memory trace, it is unclear how long the induced memory reactivation could last. To deal with this uncertainty, we modeled cue presentation with an extended boxcar-window, rather than a stick function that would restrict us to a specific time-point. The design matrix also included six regressors of no interest to account for head movement, and the three compartment regressors to account for non-specific signal fluctuations. The resulting beta estimates from the GLM were concatenated separately for cue sound and control sound trials. This procedure resulted in two separate beta time-series (i.e., consisting of 50 data points for cue sounds and 25 data points for control sounds) for all voxels in the brain. Next, voxel-wise correlation coefficients of beta time-series were computed to quantify pairwise functional connectivity for each condition separately (i.e., cue sounds and control sounds). These voxel-wise connectivity matrices were thresholded by preserving only significant connections ($p < 0.05$ False discovery rate (FDR) corrected) and setting all other connections, as well as all negative correlations to zero.

The large range of graph theoretic parameters that can be measured can result in many degrees of freedom in the analysis. Here, we restricted the analysis a priori to one metric of interest, namely the participation coefficient (Backus, Bosch, et al., 2016; Guimera, Mossa, Turtschi, & Amaral, 2005; Power et al., 2013). The participation coefficient is argued to be the most appropriate measure of hubness

(Power et al., 2013), and quantifies for each node (i.e. voxel) the diversity of its inter-modular connections (Rubinov & Sporns, 2010), specified as the amount of connections with nodes in other modules relative to the total amount of connections. It therefore is a measure of the importance of a given node for global inter-modular integration across the brain.

Following the two-step procedure used by Power and co-workers (Power et al., 2013), modular network structure was derived based on a coarser connectivity matrix, using nodes based on the 116 anatomical regions defined by the AAL (Automated Anatomical Labeling) atlas (Backus, Bosch, et al., 2016; Tzourio-Mazoyer et al., 2002). Beta time-series were averaged across voxels within an anatomical region, and a 116x116 region-by-region connectivity matrix was constructed. After thresholding this connectivity matrix (edges > 0, $p < 0.05$ FDR corrected), the matrix was parcellated into subnetworks using modularity detection according to the Louvain method (Blondel et al., 2008). Of note, the AAL-based connectivity matrix was only used to estimate the global community structure using community detection algorithms. This procedure is computationally intensive at the voxel level, but is computationally tractable using a down-sampled dataset on the basis of an anatomical parcellation (Power et al., 2013; Rubinov & Sporns, 2010).

Using the community structure estimated at the AAL-level, the participation coefficient itself was then calculated at the voxel-level, thereby retaining the fine-grained specificity at this step of analysis. Specifically, each voxel in the whole-brain voxel-wise connectivity matrix was assigned to a module, allowing the calculation of their participation coefficient ranging from 0 (provincial hub: connections are only present within its module) to 1 (connector hub: connections are only present with other modules). The resulting participation coefficient images for cue versus control sounds were compared across subjects using an undirected paired t-test.

**Seed-based connectivity analysis during cueing**

The previous analysis determined which region increased in global inter-modular network integration. We performed a follow-up seed-based functional connectivity

analysis to determine which parts of the brain were more connected with the region found during the presentation of cue versus control sounds. Therefore, we correlated the average beta time-series across all voxels of the seed region found in the network analysis (cluster-forming threshold at $Z > 2.33$, the resulting cluster was significant when testing for multiple comparisons for the whole brain using Gaussian Random Field Theory and a threshold of $p < 0.05$) with voxels in the rest of the brain separately for cue and control beta time-series. The resulting correlation maps were subsequently contrasted. Small-volume correction was used for regions that were strongly suggested to be involved on the basis of the earlier report of this data and the literature on sleep-related memory consolidation. First, the parahippocampal gyrus is activated by the object-location retrieval task (van Dongen, Takashima, et al., 2012), and is implicated in encoding visuo-spatial memories (Aguirre, Detre, Alsop, & D'Esposito, 1996; Epstein, Harris, Stanley, & Kanwisher, 1999; Epstein & Kanwisher, 1998). Furthermore, cortical regions that are involved in initial encoding and immediate retrieval prior to sleep are presumed to be part of the reactivated memory trace along with the hippocampus as a connecting hub (Danker & Anderson, 2010; Lars Nyberg, Habib, McIntosh, & Tulving, 2000; Polyn, Natu, Cohen, & Norman, 2005). Second, the thalamus and hippocampus display cue-selective activity related to overnight memory stabilization (van Dongen, Takashima, et al., 2012) and have both been implicated in sleep-related memory reprocessing and replay-related spindle-ripple events during slow-wave sleep that facilitate information transfer (Bergmann et al., 2012; Mölle et al., 2002; Skaggs & McNaughton, 1996; Staresina et al., 2015). Anatomical masks as implemented in the IBASPM 71 Atlas of the WFU Pick Atlas Tool in SPM were used for small-volume correction (Aleman-Gomez, Melie-Garcia, & Valdes-Hernandez, 2006). For statistical inferences, cluster-forming thresholds at $Z > 2.33$ were used, and the resulting clusters were tested using multiple comparisons for the whole brain using Gaussian Random Field Theory. Next, multiple-comparisons correction on anatomical regions of interest analyses was performed on the obtained test-statistics. The false-discovery rate was controlled using the Benjamini & Hochberg procedure (Benjamini & Hochberg, 1995) to calculate the critical p-value ($p = 0.031$ for this contrast) for all ROIs.

**Activity analysis during pre- and post-sleep test**

We next assessed whether network integration of visual cortex during cueing in slow-wave sleep results in the increased involvement of cortical associative regions at retrieval. The parahippocampal gyrus forms a candidate region for increased involvement with consolidation, as it is activated by the object-retrieval task, displays increased connectivity with the early visual cortex during cueing, and is generally implicated in processing visuospatial memories (Aguirre et al., 1996; Epstein et al., 1999; Epstein & Kanwisher, 1998; Peigneux et al., 2004). Thus, the parahippocampal gyrus is presumed to be the cortical memory hub that would show greater involvement with greater systems consolidation, spurred by cue-induced network participation of the early visual cortex. For this reason, we contrasted the neural activation observed during the pre-sleep (Test 1) and post-sleep (Test 2) tests, and related this contrast to the participation coefficient of the occipital cortex as observed in the graph theory analysis (see Results section). The fMRI data for Test 1 and Test 2 were analyzed in one GLM. This model included regressors for cued and uncued associations, which were modelled as delta functions and convolved with the canonical hemodynamic response function (HRF), along with temporal derivatives provided by SPM8. The design matrix included six regressors of no interest per test session to account for head movement using the realignment parameters, as well as compartment regressors accounting for non-specific signal fluctuations. Furthermore, a high-pass filter with a cutoff period of 128s was implemented to remove low-frequency fluctuations from the time series. Parameter images were generated on the first level (for each individual participant) based on a contrast between all trials of Test 1 and all trials of Test 2, to assess changes in retrieval activity from the pre- to the post-sleep test. We collapsed across cued associations and uncued associations because increased network participation of the occipital cortex in response to the presentation of cue sounds was not specifically related to only the cued associations in the learned set (see 'Results'). These contrast images were entered into a second-level GLM with one covariate, namely the participation coefficient of the occipital cortex, as found during cue sound (sounds that were paired with previously learned object-location associations) versus control sound (the sounds that were not paired with previously learned object-location associations) presentation in the sleep phase. Small-volume correction was used for the bilateral parahippocampal gyri. Here,

an anatomical mask of the parahippocampal gyrus as implemented in the IBASPM 71 Atlas of the WFU Pick Atlas Tool in SPM was used for small-volume correction (Aleman-Gomez et al., 2006). For statistical inferences, cluster-forming thresholds at Z > 2.33 were used, and the resulting clusters were tested using multiple comparisons for the whole brain using Gaussian Random Field Theory. Multiple-comparisons correction on anatomical regions of interests was performed on the test-statistics. The false-discovery rate was controlled using the Benjamini & Hochberg procedure (Benjamini & Hochberg, 1995) to calculate the critical p-value (p = 0.038 for this analysis) for each ROI.

## Results

### Network changes following cueing

We probed changes in network dynamics in response to the presentation of cue sounds versus control sounds during slow-wave sleep. The presentation of cue sounds was hypothesized to induce reactivation of the corresponding memory trace, which would presumably be reflected in increased integration of areas that represent sensory features of the cued memory representation, e.g. occipital cortex. Indeed, we found a specific increase in the participation coefficient of early visual (occipital) cortex (peak MNI coordinates: $x,y,z$ = [4, -70, 6], $z_{21}$ = 3.12, cluster-forming threshold at Z>2.33, corrected for the whole brain at p < 0.05 using Gaussian Random Field Theory), indicating that occipital cortex displays an increased inter-modular connectivity when cue sounds were presented during deep sleep (see Figure 5.2A and Table 5.1). This region exhibited overlap with regions that were activated in response to the object-location task during Test 1 (see Figure 5.2B). This overlap is suggestive of a reprocessing of the original visuo-spatial neural memory trace in response to auditory cues presented during sleep, despite the absence of any visual input.

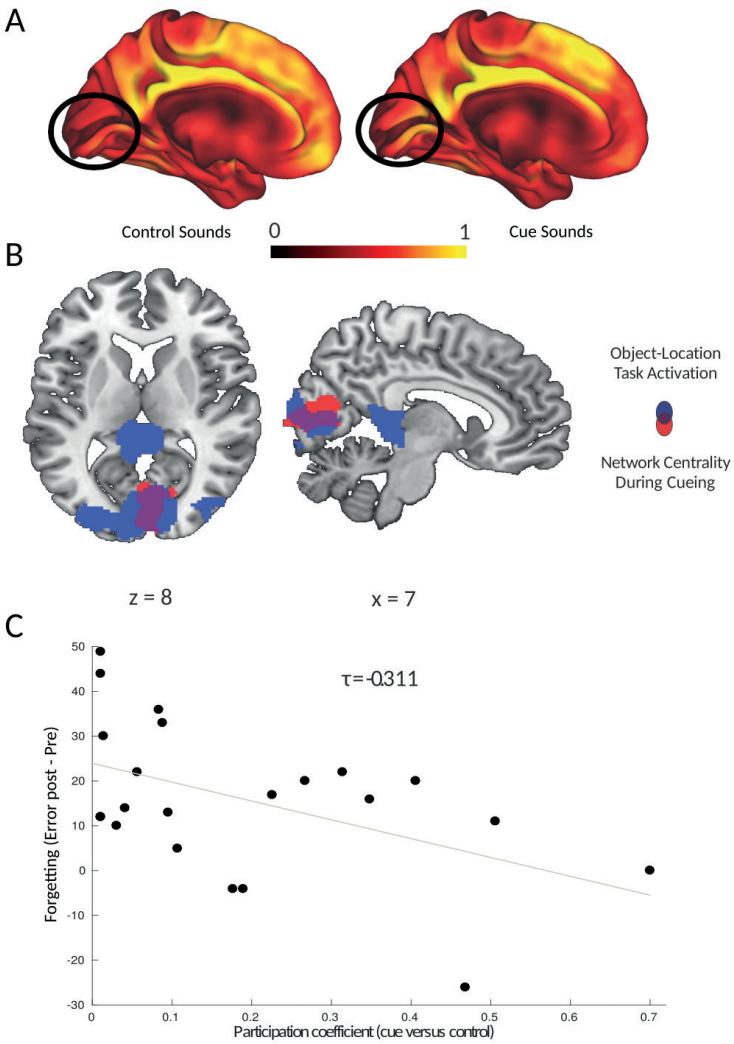



**Figure 5.2. Reactivation-related changes in network integration predict memory stabilization.**

*A. Participation Coefficient mapped for every voxel in the brain during the presentation of control sounds (sounds not previously paired with object-location information, left panel) and cue sounds (sounds previously paired with object-location associations, right panel). B. Increased network integration was found only in the occipital cortex during the presentation of cue sounds versus control sounds during slow-wave sleep. This region overlapped with the set of regions activated in response to the object-location association task. Parametric maps were superimposed onto a template brain, using the cluster-forming threshold of Z>2.33. C. The increase in participation coefficient of the occipital cortex during cueing predicted memory stabilization, indicated by reduced forgetting (expressed as the difference in error distances) between test 1 (pre) and test 2 (post).*

**Table 5.1.**

| # | Region | X | Y | Z | Peak Z-value | Cluster Size | P-value |
|---|---|---|---|---|---|---|---|
| 1 | Lingual Gyrus | 4 | -70 | 6 | 3.12 | 464 | 0.03** |
| | Cuneus | 4 | -82 | 8 | 3.06 | | |
| | Cuneus | 4 | -72 | 14 | 3.04 | | |

*Brain regions where the participation coefficient is higher during the presentation of cue sounds (previously paired with object-location associations, see right panel of Figure 5.2a), versus control sounds (not previously paired with object-location information, see left panel of Figure 5.2a). Listed are the local maxima of the significant cluster and the corresponding peak z-value. ** Region was significant using multiple comparison correction for the whole brain using Gaussian Random Field Theory and a threshold p < 0.05.*

To determine what brain regions this area in the occipital cortex preferentially connects to when recruited in the network by the auditory cues, we performed a seed-based functional connectivity analysis based on the beta time-series. The seed was defined as the region in the cortex that displayed an increase in participation coefficient during cueing (cluster-forming threshold at $z > 2.33$). The task contrast was formed by the presentation of cue sounds versus control sounds. The occipital cortex was found to be connected to the left hippocampus (peak MNI coordinates: x,y,z = [-32, -16, -18], $z_{21} = 3.36$), right hippocampus (peak MNI coordinates: x,y,z = [22, -2, -22], $z_{21} = 2.98$), left parahippocampal gyrus (peak MNI coordinates: x,y,z = [-26, -22, -26], $z_{21} = 3.10$), left thalamus (peak MNI coordinates: x,y,z = [-4, -2, -0], $z_{21} = 3.11$), right thalamus (peak MNI coordinates: x,y,z = [2, -16, -12], $z_{21} = 2.67$), and a medial prefrontal region extending from premotor cortex to anterior cingulate cortex and into dorsomedial prefrontal cortex (see Figure 5.3 and Table 5.2, all cluster-forming thresholds at $z > 2.33$, cluster-size corrected for the whole brain, or reduced search regions based on predefined anatomical areas for the bilateral hippocampus, parahippocampal gyrus and thalamus, thresholded at $p < 0.05$ using Gaussian Random Field Theory). Thus, during cueing, the occipital cortex was connected to several regions where memory replay has been shown to occur (the hippocampus, medial prefrontal cortex), and that are involved in oscillatory patterns of activity (spindles, ripples, slow-waves) that allow for systematic interactions between brain regions (e.g., the hippocampus, medial prefrontal cortex and thalamus).

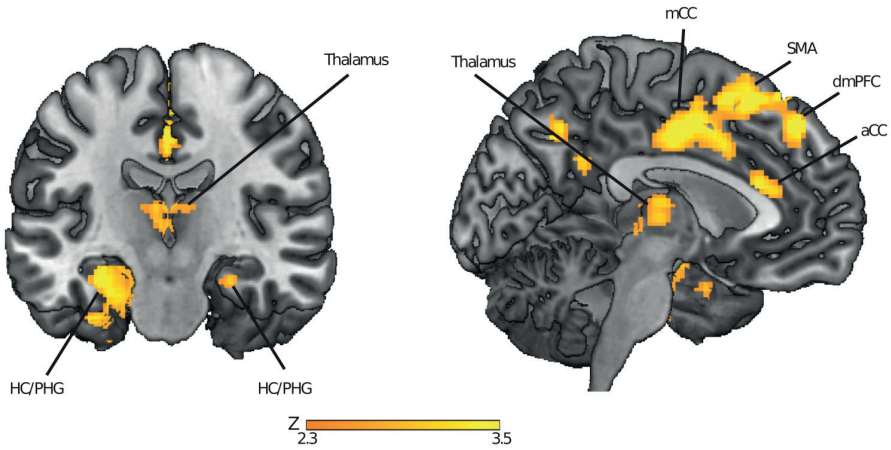



**Figure 5.3. Regions displaying greater coupling with occipital cortex during cueing in slow-wave sleep.**

*Regions displaying stronger coupling with the occipital cortex for cue sounds compared to control sounds. Parametric maps were superimposed onto a template brain, using the cluster-forming threshold of Z>2.33. HC = Hippocampus, PHG = Parahippocampal Gyrus, mCC = Middle Cingulate Cortex, SMA = Supplementary Motor Area, dmPFC = Dorsomedial Prefrontal Cortex.*

### Network dynamics related to behavioral cueing effect

Memory performance decreased on average (Test 1 performance: error = 2.74 ± 0.12 cm; Test 2 performance: error = 3.12 ± 0.14: Δerror = 0.37 ± 0.08; $t_{21}$ = 4.48, p < 0.001), both for cued (Δerror = -0.44 ± 0.11 cm, $t_{21}$ = 3.86, p = 0.001), and uncued associations (Δerror = −0.33 ± 0.11 cm, $t_{21}$ = 2.56, p = 0.018). There was no effect of cueing on participants' memory accuracy as tested after sleep ($F_{1,21}$ = 0.83; p = 0.374). Thus, there was forgetting across the sleep/rest session, and this forgetting appeared to be similar for cued and uncued object-location associations.

**Table 5.2.**

| # | Region | X | Y | Z | Peak Z-value | Cluster Size | P-value |
|---|---|---|---|---|---|---|---|
| 1 | Anterior cingulate gyrus | 16 | 14 | 36 | 3.75 | 2218 | 0.003** |
| | dorsomedial prefrontal cortex | -10 | 30 | 54 | 3.65 | | |
| | dorsomedial prefrontal cortex | -10 | 36 | 56 | 3.5 | | |
| | rostral cingulate gyrus | -2 | -6 | 42 | 3.38 | | |
| | dorsomedial prefrontal cortex | -6 | 44 | 44 | 3.35 | | |
| | Supplementary motor area | 0 | 2 | 44 | 3.35 | | |
| 2 | L hippocampus | -32 | -16 | -18 | 3.36 | 271 | 0.006* |
| 3 | R hippocampus | 26 | -18 | -18 | 2.69 | 13 | 0.031* |
| 4 | L parahippocampal gyrus | -26 | -22 | -26 | 3.1 | 180 | 0.010* |
| 5 | L thalamus | -6 | -16 | 12 | 2.81 | 215 | 0.013* |
| 6 | R thalamus | 2 | -16 | 12 | 2.67 | 74 | 0.016* |

*Brain regions that show a significantly higher connectivity with the early visual cortex during the presentation of cue sounds versus control sounds. Listed are the local maxima of the significant cluster, as well as clusters found in pre-defined anatomical regions of interest. The critical p-value was calculated in order to correct for multiple (six) comparisons performed using anatomical regions of interest using the Benjamini & Hochberg procedure (1995) to control the false discovery rate. The critical p-value thus calculated was p = 0.031.*
*** Region that was significant using multiple comparison correction for the whole brain using Gaussian Random Field Theory and a threshold p < 0.05.*
** Region that was significant using a multiple comparison correction for a reduced search volume defined by anatomical regions of interest, and a subsequent correction for the false discovery rate (critical p-value = 0.031).*

However, there was substantial variability in the effect of cueing on retention. Therefore, different neural responses to cueing might explain individual differences in the effect of cueing. In particular, increased network participation of occipital cortex could be an index of how much the visual-spatial memory trace became reactivated in a given subject when auditory cues were presented. Indeed, the increased participation coefficient during cue sound presentation was found to be associated (see Figure 5.2) with reduced overnight forgetting of visual-spatial object locations ($\tau_{21}$ = -0.311, p = 0.048). This relationship was specifically present

for the visual-spatial locations that were associated with sound cues ($\tau_{21}$ = -0.338, p = 0.032), but not for other locations ($\tau_{21}$ = -0.211, p = 0.183), although the difference in these correlations did not reach significance (Williams-Hotelling test, t = 0.43, p = 0.668). Furthermore, the increased participation coefficient during cue sound presentation was specifically related to overnight forgetting, and not to general memory performance as assessed at Test 1 ($\tau_{21}$ = 0.048, p = 0.777) or Test 2 ($\tau_{21}$ = -0.092, p = 0.572). This result indicates that those participants who displayed increased network participation of occipital cortex experienced a reduction in overnight forgetting of all associations. This effect was not specific to the cued items, but extended to uncued object-location associations encoded within the same spatial layout and temporal learning context.

**Retrieval activity changes related to network integration during cueing**

We then considered whether the increased network integration of the occipital cortex induced by cueing during sleep had subsequent effects on neural activity when retrieving object-locations during the post-sleep test. Here, we found that the participation coefficient was related to a higher retrieval-related activity in Test 2 relative to Test 1 in both left parahippocampal gyrus (peak MNI coordinates: *x,y,z* = [-16, -32, -18], $z_{21}$ = 4.14), and right parahippocampal gyrus (peak MNI coordinates: *x,y,z* = [16, -32, -18], $z_{21}$ = 3.26, cluster-forming threshold at Z>2.33, corrected for a reduced search volume based on anatomical masks of left and right parahippocampal gyrus at a threshold of p < 0.05 using Gaussian Random Field Theory). Thus, increased network integration of occipital cortex during cueing in sleep had a subsequent effect on the neural regions that were activated during retrieval of object-locations at Test 2, thereby recruiting bilateral parahippocampal gyrus at Test 2 to a relatively greater extent (see Figure 5.4 and Table 5.3).

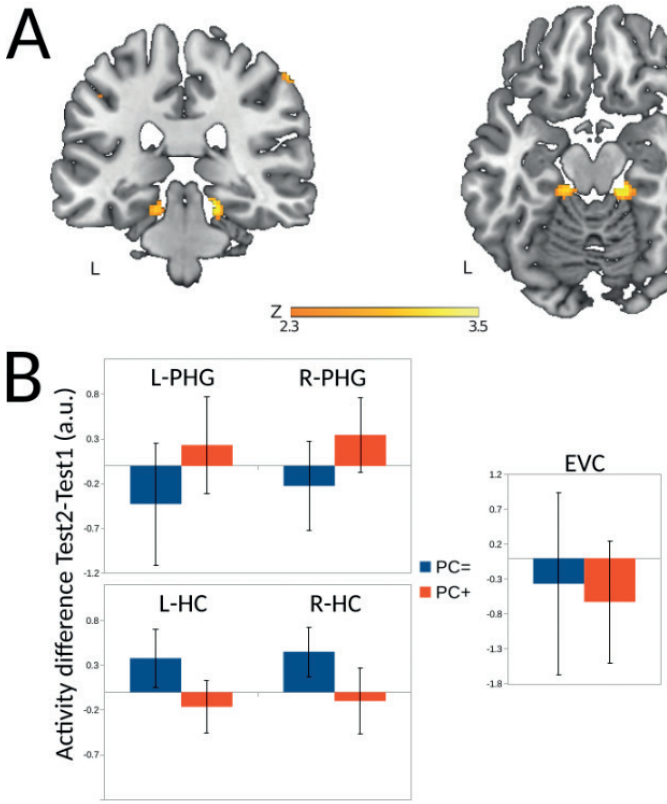



**Figure 5.4. Overnight activity increases related to network integration of occipital cortex.**

*A. Regions in the parahippocampal gyrus where there is a relation, on the one hand, between an activity increase between the pre-sleep and post-sleep test and, on the other hand, increased network integration of occipital cortex during cueing in the slow-wave sleep period between Test1 and Test 2. Parametric maps were superimposed onto a template brain, using the cluster-forming threshold of Z>2.33. B. For comparison purposes, the mean overnight regional activation differences (calculated as activity during post-sleep Test 2 minus pre-sleep Test 1) are plotted for the left (L-PHG) and right (R-PHG) parahippocampal gyrus, the left (L-HC) and right (R-HC) hippocampus, and early visual cortex (EVC) for two groups of participants based on a median split of the participation coefficient scores in the early visual cortex. Even though there were no significant differences here across the whole region, the numerical values are broadly in agreement with our hypotheses, namely finding a reduced activity of the hippocampus at Test 2 and an increased activity of the parahippocampal gyrus for the PC+ group (those participants that did display the strongest increase of the participation coefficient) versus the PC= group (those participants that displayed the lowest increase of the participation coefficient), while there were no visible differences in the early visual cortex.*

**Table 5.3.**

| # | Region | X | Y | Z | Peak Z-value | Cluster Size | P-value |
|---|---|---|---|---|---|---|---|
| 1 | L parahippocampal gyrus | -16 | -32 | -18 | 3.26 | 40 | 0.038* |
| 2 | R parahippocampal gyrus | 16 | -32 | -18 | 2.81 | 215 | 0.013* |

*Brain regions that show a significant relation between an increase in network integration of occipital cortex during cueing in slow-wave sleep and activity on the post-sleep test versus the pre-sleep test. Listed are clusters found in pre-defined anatomical regions of interest. The critical p-value was calculated to correct for multiple (2) comparisons using anatomical regions of interest using the Benjamin & Hochberg procedure (1995) to control the false-discovery rate. The critical p-value thus calculated was p = 0.038.*
** Region was significant using a multiple comparison correction for a reduced search volume defined by anatomical regions of interest, and a subsequent correction for the false discovery rate (critical p-value = 0.038).*

## Discussion

Here, we investigated the whole-brain neural network reorganization induced by cued reactivation of memory traces during deep sleep, and its relation to behavioral and neural responses during subsequent retrieval of those memories. For that purpose, we re-analysed fMRI data reported in the study by van Dongen and colleagues (van Dongen, Takashima, et al., 2012) using whole-brain connectivity measures. We demonstrated that when participants were in a state of slow-wave sleep, absent of visual input, and were then cued with sounds associated with previously learned visuospatial information, the occipital cortex displayed an increased network integration as measured with the participation coefficient. Furthermore, subjects who displayed increased cue-induced network integration of occipital cortex showed increased memory stabilization. This finding is congruent with the notion that the sound cue induces global reprocessing of the learning episode that may be conducive to subsequent memory retention. Although the occipital cortex is not typically a focal point for mnemonic processing, previous studies have shown that cue-triggered reactivation can lead to instantiations of previously experienced episodes in this region (Ekman, Kok, & de Lange, 2017; S. Xu, Jiang, Poo, & Dan, 2012). It is likely that this reinstatement is coordinated with other mnemonic regions. Indeed, various regions displayed increased connectivity with the occipital cortex during the presentation of cue sounds: memory and

replay-related regions, such as the hippocampus, thalamus and medial prefrontal cortex, and higher-order associative cortices implicated in the encoding of the memory trace, such as the parahippocampal gyrus. These results are in line with a study (Ji & Wilson, 2007) that showed coordinated replay between the hippocampus and occipital cortex, the perceptual region that was involved in the initial processing of the stimulus. Cueing during slow-wave sleep also showed an enduring relation to neural activation at memory retrieval as measured during the post-sleep retrieval test. Specifically, subjects with greater network integration of the occipital cortex during slow-wave sleep recruited parahippocampal gyrus more at subsequent retrieval relative to pre-sleep retrieval.

First, it should be reiterated that auditory cueing did not have an overall effect on memory retention of cued versus uncued associations. Similarly, the network reorganization found during the presentation of cue sounds was not specifically related to memory stabilization of those object-location memories associated with the cues. The association of network participation with memory stabilization was numerically larger for cued versus uncued material, but the interaction was not statistically significant. However, there could be various explanations for a nonspecific effect across all associations. First, the MRI scanning environment is extremely loud, unlike the quiet environment that would be preferred for auditory stimulation during sleep. Subjects who succeed in falling asleep were also the ones who succeeded in filtering out sensory input the best. Therefore, it is conceivable that the acoustic stimuli effectively modulated ongoing brain processes in some individuals but not others, contributing to large inter-individual differences and lower power of demonstrating this effect as a group than in studies conducted outside of the MR-scanner environment (Rudoy et al., 2009). It is also possible that the study was underpowered to observe an interaction effect, and that a larger sample would have rendered a significant interaction of cueing and network participation of the early visual cortex on overnight memory change. Furthermore, it is possible that the cueing had a non-specific effect on all material acquired within the same temporal and spatial learning context (as also suggested by the results reported in Oudiette et al., 2013; Rasch et al., 2007). Indeed, all object-location associations were acquired within the same learning context; the same scanning session and temporal learning sequence, and most importantly, all objects

were located on the same two-dimensional spatial grid. Therefore, it is plausible that the various object-locations became integrated into a unitary map, or spatial schema (Tse et al., 2007; van Buuren et al., 2014) at initial learning and/or during replay (Durrant, Cairney, McDermott, & Lewis, 2015; Hennies, Ralph, Kempkes, Cousins, & Lewis, 2016). As such, object locations could have been stored not only with reference to the two-dimensional grid, but also directly or indirectly in spatial reference to the other objects. As such, a sound cue could have reactivated cue-sound specific object-location associations, but also objects located in the objects' vicinity. In fact, it has been proposed that sleep-dependent memory replay is instrumental in the integration of memories into a cognitive schema (Lewis & Durrant, 2011). Thus, while network integration of the occipital cortex during cueing is demonstrated to be positively related to overall memory stabilization of a set of object-location associations, the fact that an integrated spatial schema was cued could have contributed to non-specific reactivation of all object-locations within this set. Previous studies might have been powerful enough to find a cue-selective effect of targeted memory reactivation (Creery et al., 2015; Rudoy et al., 2009), over and above this context-general effect. Future studies using targeted memory reactivation should control for this by testing uncued material learned in the same context as the cued material, along with uncued material acquired in a different learning context, in order to disentangle these effects.

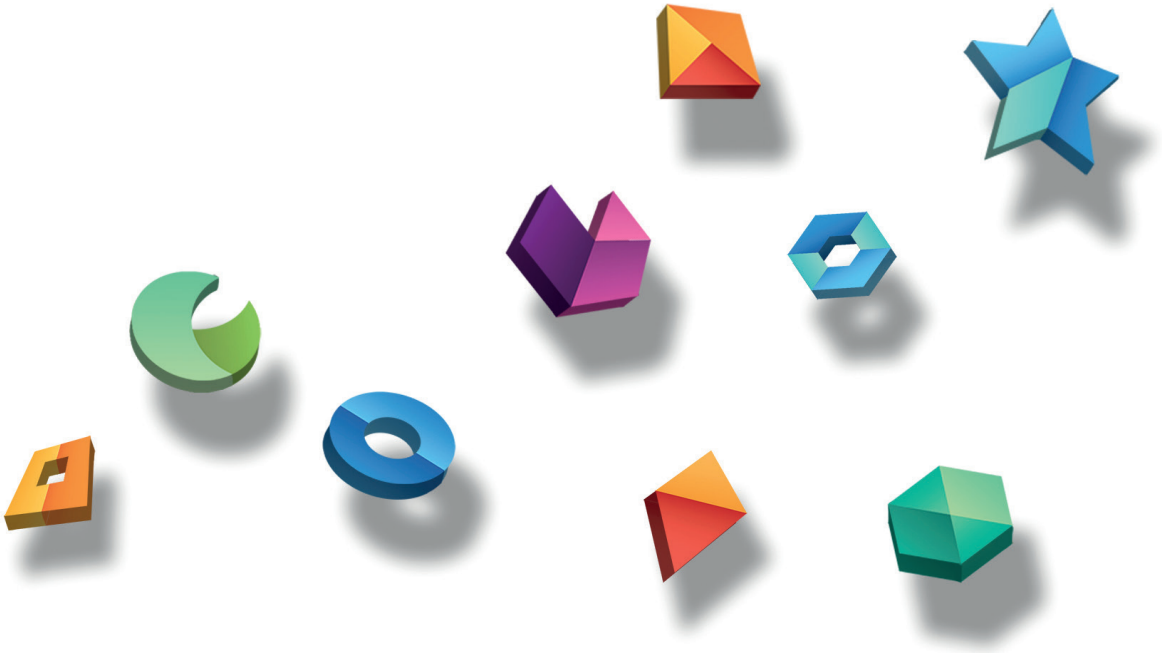
An alternative explanation of the brain-behavior relationship can be proffered too. It could be that for those subjects who display better memory consolidation after sleep, re-exposure of the cues during sleep induces a better memory reactivation. Here, a stronger association strength between the cue and the memory trace during slow-wave sleep prior to cueing could result in stronger memory reactivation following cue presentation. This alternative explanation, where reactivation does not improve memory but rather is a result of better memory, cannot be ruled out based on the current data.Bearing in mind these caveats, our results show a network reorganization in response to cues presented during slow-wave sleep that is consistent with an active model of sleep-dependent memory consolidation. Previous research has shown that whole-brain connectivity patterns display a general reduction in thalamo-cortical and cortico-cortical connectivity and an increase in local clustering during slow-wave sleep (Spoormaker et al., 2010).

Against this backdrop, a hippocampal-neocortical dialogue takes place facilitating systems-level consolidation by integrating hippocampal-dependent memories into neocortical storage sites (Buzsáki, 1996; Mitra et al., 2016). This hippocampal-neocortical dialogue is orchestrated through oscillatory electrophysiological patterns characteristic of slow-wave sleep and related to memory replay (Davidson, Kloosterman, & Wilson, 2009; Diba & Buzsáki, 2007). Specifically, slow (delta wave) oscillations propagate across the brain and to the medial temporal lobes, including the hippocampus, exerting a global control over spiking activity (Massimini, Huber, Ferrarelli, Hill, & Tononi, 2004; Nir et al., 2011). Mechanistically, the up-state of slow oscillations has been found to enable thalamic sleep spindles in the sigma band, which in turn cluster high-frequency hippocampal ripples in their troughs. These hippocampal ripples then are proposed to allow for a precisely timed exchange of mnemonic information between the hippocampus and neocortex, potentiating and strengthening the neocortical memory trace. If external auditory cueing would induce a reactivation of the visuo-spatial memory traces, one would expect that cortical representation areas corresponding to the modality of the memory trace would be recruited to participate in this cross-talk. Indeed, the occipital cortex displayed an increase in global network integration during cueing, and when probing the neural connectivity changes of the occipital cortex that was paired with this increase, we found that this region increases its connectivity with the regions involved in active cross-talk during slow-wave sleep, namely the hippocampus, thalamus and medial prefrontal cortex (Mölle, Yeshenko, Marshall, Sara, & Born, 2006; Staresina et al., 2015). Of note, our findings do not exclude the possibility that connectivity within the visual cortex is also increased in response to cueing. The findings merely indicate that inter-modular connectivity is preferentially increased. A reinstated memory trace consists of connections between lower-level perceptual regions such as the early visual cortex (representing low-level features) and higher-level associative/mnemonic regions such as the hippocampus and parahippocampal gyrus (which bind together low-level features; Nadel et al., 2000; Squire & Alvarez, 1995). Moreover, memory reactivation during sleep is thought to be paired with a synchronization of neural activity across disparate brain regions, facilitating the exchange of information proposed to be necessary for systems-level memory consolidation (Bergmann et al., 2012; Marshall & Born, 2007; Staresina et al., 2015). Therefore, an increase in inter-

modular connectivity of regions representing features of the memory trace reflects processes that are expected to be induced by memory cueing.It is plausible that the neural results reported here are akin to those found during endogenous memory replay that are predominantly occurring during slow-wave sleep. Certainly, the neural patterns reported here match presumed neural mechanisms of memory replay during slow-wave sleep. However, the same neural pattern may also be found in response to memory reactivation during other brain states, such as wake, light sleep or REM-sleep. Based on the current data, the extent to which the neural memory reactivation pattern described is specific to slow-wave sleep cannot be assessed, as cues were only presented during slow-wave sleep. Future studies may compare the neural effects of cued memory reactivation across different states to determine which aspects are specific to slow-wave sleep and which are state-independent.At first sight, it might be surprising that the auditory cortex was not recruited in response to auditory cueing. However, it should be noted that a contrast was made between two conditions in which sounds were presented. The critical difference in the contrast was that cue sounds but not control sounds were associated with specific visuo-spatial memories. Given the multimodal learning conditions, one could still expect increased connectivity of the auditory cortex with visual, spatial, and memory regions during cueing. As the functional connectivity analyses were informed by results of the graph theory analyses, we did not further probe connectivity from an anatomically defined seed in auditory cortex. Despite this caveat, it remains to be noted that cueing recruited occipital cortex to engage with mnemonic regions that have been shown to display increased cross-talk during slow-wave sleep, and this was related to the stabilization of the associated memory traces in neocortical storage sites.Sleep supports memory retention in a selective manner, and information is prioritized based on perceived future relevance at encoding (van Dongen, Thielen, Takashima, Barth, & Fernández, 2012; Wilhelm et al., 2011) emotional salience (Hu, Stylos-Allan, & Walker, 2006; Payne et al., 2007; Wagner, Hallschmid, Rasch, & Born, 2006) and consistency across episodes and with prior knowledge (Durrant et al., 2015; Tamminen et al., 2013). Possibly, relevant information is selectively prioritized for active processing during sleep, whereas non-prioritized memory traces are forgotten, perhaps aided by the homeostatic regulation of synaptic plasticity through global downscaling (Genzel, Kroes, Dresler, & Battaglia, 2014; Tononi & Cirelli, 2003, 2006). The net

result is an increase in signal-to-noise for the prioritized memory traces in the neocortex, contributing to memory stabilization. Indeed, this appears to be the case in the parahippocampal gyrus, where an increase in retrieval-related activation of posterior parahippocampal gyrus at post-sleep test versus pre-sleep test was related to an increase in network participation of occipital cortex during cueing in sleep. The parahippocampal gyrus is known to be involved in representing the local visual environment (Epstein & Kanwisher, 1998) and learning spatial layouts (Aguirre et al., 1996). The increased involvement of parahippocampal gyrus during post-sleep retrieval is consistent with a role of parahippocampal gyrus in binding together low-level visual information from downstream occipital cortex. Notably, parahippocampal gyrus was activated during retrieval at the pre-sleep test, showed selective increased activation in response to cueing, and displayed increased connectivity with occipital cortex during cued reactivation, as originally reported (van Dongen, Takashima, et al., 2012).

In sum, here we showed that inducing reactivation of object-location memories with associated auditory cues during slow-wave sleep increased the integration of information in the occipital cortex within a global brain network, specifically including those regions involved in memory replay, such as the medial temporal lobe, thalamus and medial prefrontal cortex. Furthermore, global network integration of the occipital cortex during cueing showed a correlation with overnight memory stabilization of learned materials and involvement of the parahippocampal gyrus during post-sleep retrieval. These findings highlight how graph theory analysis can be used to assess whole-brain connectivity patterns during targeted memory reactivation in slow-wave sleep, and contribute to a better understanding of sleep-related memory consolidation.



# CHAPTER 6

General discussion

In the **Introduction**, I outlined a putative framework, based on prior experiments and theoretical models, of how the human brain selectively integrates information from ongoing experience into long-term memory storage. This framework informed the research questions regarding memory encoding and consolidation:

**Q1: How are inter-individual differences in emotional memory explained by activity and connectivity during encoding?**

**Q2: What influence does perturbation of the medial prefrontal cortex have on schema-related memory errors?**

**Q3: What brain regions are involved in the accumulation and updating of novel generalized knowledge structures?**

**Q4: What brain connectivity changes are induced by cued memory reactivation during sleep and are they related to memory stabilization?**

In **Chapters 2-5**, I then described four experimental studies that aimed to answer these questions, by considering emotional memory encoding, schemas and false memories, and the acquisition of generalized knowledge, and lastly sleep and memory reprocessing.

In this Chapter, I will summarize how the experimental findings reported in the previous chapters contribute to formulating an answer to these research questions. Moreover, the reported experimental results and their implication may motivate an updating of the theoretical framework. Specifically, I will discuss how the results reported here, help to answer the central research question: ***How does the human brain selectively integrate information from incoming experience into the long-term memory store?*** It is important to note that reported findings and proposed models are tentative and prone to further replication and refinement through added experimentation, data, and improved models. This iterative process is the bedrock of science and scientific progress. As such, these experimental studies should not be considered in isolation, and be given meaning in the larger body of literature and experimental findings. Thus, I will aim to give the reported

results this context and suggest directions for further inquiry to develop our model of human memory and the brain. An in-depth discussion of the individual experiments, and both their merits and limitations is already given in the individual experimental chapters (**2-5**). I will therefore focus here on how these experiments contribute to understanding how our brain selectively integrates information from incoming experience into the long-term memory store, and why this mechanism is an adaptive feature of the human brain.

## Emotional memory encoding

In the study described in **Chapter 2**, I attempted to answer the question: **How are inter-individual differences in emotional memory explained by activity and connectivity during encoding?** Here, I reported a paradoxical emotional impairment effect on memory specificity for the recall of face-identity associations in a large sample of participants. I measured a selective increase in activity in the medial prefrontal cortex during the encoding of associations with a negative valence (versus those with a neutral valence). Furthermore, the medial prefrontal cortex was increasingly connected with both the hippocampus and the amygdala during the encoding of negative associations. The strength of this connectivity with specifically the hippocampus was related to individual differences in the extent to which there was a loss in memory specificity. This finding is in line with a role of an mPFC-hippocampal circuit regulating memory specificity (see also Xu & Südhof, 2013).



The findings add to several lines of inquiry regarding the role of these regions in regulating memory encoding. The congruency of encoded information with an established schema has also been found to influence memory specificity, potentially by regulating hippocampal activity (van der Linden et al., 2017; van Kesteren et al., 2013, 2014). However, other accounts have suggested that connectivity between the mPFC and hippocampus actually reduces memory interference and increases pattern separation (Guise & Shapiro, 2017). Alternatively, connectivity between the mPFC and hippocampus is suggested to support memory updating, and integration across multiple episodes (Backus, Schoffelen, Szebényi, Hanslmayr, & Doeller, 2016; Schlichting & Preston, 2016). To these existing accounts we can add another characteristic: with negative valence of memory encoding, there

is an increased interplay between the hippocampus and the medial prefrontal cortex paired with a reduced memory specificity. The exact mechanism by which the medial prefrontal cortex may interact with the hippocampus, potentially mediated by the nucleus reuniens of the thalamus (Thielen et al., 2015), remains yet to be fully elucidated. These seemingly differing existing lines of inquiry on mPFC-hippocampal interactions converge in highlighting a role in the regulation of memory encoding.

The pattern of behavioral results is in line with other studies that suggest that emotional valence has a qualitatively different effect on item memory and associative memory. Whereas there is generally an emotional enhancement effect to be found on item memory, an emotional impairment effect can be found on associative memories (Bisby & Burgess, 2014). Similarly, a reduction in hippocampal activity has also been reported during the encoding of memories of a negative valence (Bisby, Horner, Hørlyck, & Burgess, 2016), and hippocampal activity at encoding was found to be predictive of associative memory, while amygdala activity at encoding was predictive of item memory. However, while this particular study did not report any involvement of the medial prefrontal cortex, it does suggest an interesting dynamic between activity in the hippocampus and amygdala, and associative memory versus item memory. It should be noted that I also report a simultaneous increase in connectivity of the mPFC with the amygdala, a region that is located adjacent to the hippocampus. Furthermore, hippocampal activity was reduced during negative associative encoding, in line with an inhibitory account of the mPFC-hippocampal circuit. It could be that during negative associative encoding, there is a shift towards gist-based memory encoding, benefiting memory of items, at the expense of associated peripheral detail, thereby impairing associative memory encoding. This suggests that in this study, recognition (item) memory for the faces and occupations in isolation might have been unimpaired or even enhanced. Future studies should compare item and associative memory encoding directly in a similar large-scale study design.

Furthermore, the mechanism by which the nodes in this circuit exert a mutual influence should be elucidated further as well. One possibility is that the amygdala signals the emotional valence to the medial prefrontal cortex, which in turn,

inhibits hippocampal encoding or shifts hippocampal encoding from its posterior aspects to its most anterior aspects. Another possibility is that the amygdala directly inhibits hippocampal encoding and upregulates the medial prefrontal cortex. A third possibility is that the medial prefrontal cortex directly inhibits hippocampal encoding. All these possibilities should be elucidated further, potentially using time-resolved electrophysiological recordings. Clarification of these neural mechanisms allows the probing of how these mechanisms differ across individuals in large samples, and potentially explain some cognitive deficits in clinical samples with deficits in emotional memory specificity (and overgeneralization of emotional memories), such as in social phobia, depression and PTSD (Brown et al., 2013; Foa et al., 2000; Watkins et al., 1996), as well as differences in emotional memory across aging (Henson et al., 2016).

Such a reduced specificity of emotional memory encoding may also have had evolutionary benefits. It might be beneficial for survival to be sensitive to potential threats and negative outcomes (van Marle et al., 2009). Therefore, it might be adaptive to overgeneralize a learned negative association, from one particular object or central feature, to other similar objects or features with slightly differing associated peripheral information. In ancient societies, it may have thus been more adaptive to be sensitive (from a signal processing perspective), in order to be vigilant towards threats. However, in modern-day society most direct threats to human survival have been eradicated, and as such an overgeneralization of emotional responses may well result in cognition and behavior that we consider maladaptive, such as in depression and anxiety disorders (Kheirbek, Klemenhagen, Sahay, & Hen, 2012).



## Schemas and false memories

In **Chapter 3** I set out to answer the question: **What influence does perturbation of the medial prefrontal cortex have on schema-related memory errors?** Ongoing processing in the medial prefrontal cortex was perturbed for a sustained offline period using transcranial magnetic stimulation (specifically a repetitive TMS-protocol, called continuous theta-burst stimulation). When participants subsequently performed the DRM-task, a reduction in false recall of critical lures was observed, compared to a stimulation control group and a behavioral control

group. True recall and recognition memory performance (false or true) remained unaffected across groups. These results mirror the findings of another study that found that damage to the medial prefrontal cortex reduced false recall of critical lures, using a similar task design in patients with focal brain lesions (Warren et al., 2014). This data provides initial causal evidence for a role of the mPFC in the encoding and subsequent retrieval of false memories. These false memories were induced by design through the induction of strong expectations of the critical lure, on the basis of prior knowledge or schemas. By down-regulating the functioning of the mPFC, the schema is presumably given less weight in encoding and/or retrieving new memories, thus resulting in reduced false recall.

Some intriguing questions remain, that should be addressed in further studies. First of all, how does the brain stimulation protocol affect underlying activity and connectivity in the underlying brain regions? The notion that locally applied TMS may affect not only the underlying brain region, but also downstream interconnected regions is certainly bolstered by combined TMS-fMRI studies (Eldaief, Halko, Buckner, & Pascual-Leone, 2011; Fox, Halko, Eldaief, & Pascual-Leone, 2012; Jung, Bungert, Bowtell, & Jackson, 2016), and forms an avenue to pursue in future research. Second, the study reported here employed a design where various lists of words were encoded and immediately recalled. Therefore, the present data would be insufficient to be able to disentangle effects on encoding or immediate recall. Future studies should address this by temporally separating the encoding, recall and recognition in time. In such a design, stimulation should either occur immediately prior to encoding, upon which there is a sufficient time period to let the effects of the stimulation dissipate before recall and recognition are tested, or following encoding but preceding recall. A third factor to be considered is that qualitative sleep parameters in the night prior to the experiment affected mnemonic task performance, not only in terms of correct recall but also the false recall of critical intrusions. The finding that the perceived sleep quality in the night before encoding positively influences false memory errors is intriguing. Most of all, these findings highlight the fact that these factors need to be taken into account when assessing performance on the DRM-task, and that the exact manner in which sleep prior to task performance influences the DRM-task need to be elucidated in detail further in future studies.

Even though in this particular situation, the availability of prior knowledge may lead to increased memory errors, it may be an adaptive feature of our memory system that we encode information differently depending on whether it is consistent with prior knowledge. For instance, simulations using biologically plausible neural network models have demonstrated that the rapid integration of consistent information does not interfere with prior established knowledge (McClelland, 2013). However, when inconsistent information is rapidly integrated, catastrophic interference takes place with established knowledge representations. Thus, for new incoming information that is inconsistent, it is necessary to encode information using the hippocampus to prevent catastrophic interference. The added but crucial benefit of hippocampal encoding is that an event is encoded in a rich, detailed manner, including its spatiotemporal context. This allows the brain to better predict these novel, inconsistent events in the future. In contrast, most ongoing experiences are broadly consistent with our schemas, and for this purpose, we can encode them in a different, more efficient manner by rapidly integrating relevant aspects into the existing memory store without encoding unnecessary detail and spatiotemporal context. The flipside of this different mode of encoding, potentially mediated by the mPFC, is that strong schema-based expectations about certain events may actually foster erroneous memories of their occurrence.



## Acquisition of generalized knowledge

In **Chapter 4,** I posed the question: **What brain regions are involved in the accumulation and updating of novel generalized knowledge structures?** In this study, I systematically tracked knowledge formation whilst participants were learning an abstract artificial language organized by higher-order associative regularity. By explicitly modelling the state of knowledge at each trial, I was able to carefully track the accumulation and updating of knowledge during feedback-based learning and subsequent cue-based retrieval. I found that activity in the left inferior frontal gyrus (left IFG) increases both with feedback-based knowledge updating, and available accumulated knowledge during the cue-phase. Furthermore, in the right caudate nucleus, activity was found to correlate with the availability of recently acquired knowledge during the cue-phase, suggesting it might contribute to retrieving recently updated knowledge. Furthermore, during feedback processing, I found that in a set of regions, including the medial prefrontal

cortex and posterior cingulate cortex, activity scaled with previously accumulated knowledge. This could mean that increased generalization of knowledge with increased knowledge accumulation might take place in these regions.

The knowledge updating parameter selectively revealed two brain regions, namely the left IFG and the right caudate nucleus. These regions are presumably actively involved in updating (left IFG) and recruiting recently updated knowledge (right caudate nucleus) of the associations comprising the linguistic structure based on trial-specific information. According to a recent review (Scimeca & Badre, 2012), the striatum, including the caudate nucleus, might play three roles during retrieval of declarative memories, namely modulating the re-encoding of retrieved items according to their expected utility (adaptive encoding), selecting information in working memory to aid successful retrieval (adaptive gating), and adjusting cognitive control based on retrieval outcomes (reinforcement learning). All three of these roles of the striatum potentially fit the role that the caudate nucleus plays in the current experiment. Based on feedback on the prior trial, knowledge is updated and this updated knowledge needs to be retrieved in the next trial. The more knowledge is updated, the more this knowledge needs to be taken into account when retrieving information in the next trial, and this information needs to be selected for use in the next trial (adaptive gating) and re-encoded for all subsequent trials (adaptive encoding).

While the caudate nucleus likely reflects a domain-general learning structure, the left IFG may be a region specifically involved in acquiring linguistic knowledge (Hagoort, 2005). Alternatively, the region may fulfil a more general role in the learning of sequences that conform to a certain finite-state grammar, along with striatal regions (Peigneux et al., 1999; Uddén & Bahlmann, 2012). Whereas these regions may be involved in the active updating of an acquired rule structure, other regions may be involved too. Presumably, the detection of the regularity of the rule structure may involve combining information across successive trials. The hippocampus may be involved in detecting these regularities across individual learning episodes along with the medial prefrontal cortex (Kumaran & McClelland, 2012). The knowledge updating parameter may not pick up on the precise neural dynamics of regularity detection. Regularity detection presumably takes place

already at an early stage and presumably is prerequisite for knowledge updating to be able to take place. Moreover, regularity detection presumably continues across learning as more regularities are picked up, which may then be reflected in increasing activity with accumulated knowledge during feedback presentation.

The left IFG is also increasingly active with accumulated knowledge during the retrieval phase, indicating that it scales with the recruitment of increasing amounts of knowledge. It may be that this area represents the actual knowledge pertaining to the associative linguistic structure. The exact format of the representations of the associative linguistic structure needs to be probed further in future fMRI-studies. Multivariate pattern analysis techniques such as decoding (Haynes & Rees, 2005; Kamitani & Tong, 2005), representational similarity analysis (Kriegeskorte, Mur, & Bandettini, 2008) and pattern-component modeling (Diedrichsen, Yokoi, & Arbuckle, 2018) can be leveraged to look at the representational format of the associative linguistic structure across different brain regions, specifically probing where lower-level associative rules are stored and where the integrative information beyond individual associative rules is represented. Some prior work has already been done. For instance, different components of a schematic associative structure have been shown to be decodable from the angular gyrus, which presumably might constitute a hub where these different schema components are integrated into a higher-order structure (Wagner et al., 2015). A similar analysis of data obtained using my paradigm could show whether the present linguistic structure can similarly be decoded from the angular gyrus, the medial prefrontal cortex, or alternatively, the left IFG.

A mechanism to gradually train the neocortex on generalized associative knowledge across trials, as captured here with our learning models, is an adaptive feature of our learning brain. For instance, computational models have shown that the neocortex should not immediately encode all incoming experiences, as this would lead to catastrophic interference with established knowledge (McClelland et al., 1995). Rather, our brains likely have developed two complementary learning systems. One learning system that encodes consciously apprehended individual experiences in vivid detail with spatiotemporal context (Moscovitch, 2008, but see Maguire & Mullally, 2013). A second neocortical learning system may readily

integrate consistent information (McClelland, 2013; Tse et al., 2011) and be slowly 'trained' through repeated reprocessing during sleep (see next section), or through interleaved, repeated learning experiences (McClelland et al., 1995). The work described in this chapter incorporates roles for the caudate nucleus, left inferior frontal gyrus, hippocampus and medial prefrontal cortex into this mechanism of training the neocortex across repeated, interleaved learning.

## Sleep and memory reprocessing

In **chapter 5** I aimed to answer the research question: **What brain connectivity changes are induced by cued memory reactivation during sleep and are they related to memory stabilization?** Specifically, I found that the occipital cortex, a visual region, displayed greater network integration (as measured by the participation coefficient) in response to auditory memory cues during slow-wave sleep. This region is presumably involved in representing low-level aspects of the encoded memories. These memories consisted of associations between a sound, and a visual object at a particular location on a 2D grid. The occipital region is considered to be low-level visual cortex, and moreover, it overlaps with the set of regions found to be active in response to the initial pre-sleep retrieval of the memory trace. This suggests that aspects of the original memory trace are reprocessed in response to the cue, and I showed that this reprocessing is related to overall memory stabilization. Specifically, the reported increase in network participation of the occipital cortex was related to the difference in memory performance between the post-sleep and pre-sleep test, indicating that such memory reprocessing might be beneficial for memory retention.

The standard model of memory consolidation states that an associative memory trace is initially encoded in posterior representational regions coding for low-level features, which are bound together by medial temporal lobe regions, especially the hippocampus and parahippocampal gyrus (Squire et al., 2015). The parahippocampal gyrus is a memory region that is known to be specifically involved in acquiring and representing the local visual environment (Aguirre et al., 1996; Epstein & Kanwisher, 1998). Presumably, memory traces laid down prior to sleep are reprocessed during sleep such that its features become better embedded in the neocortex (Stickgold et al., 2001). Specifically, memory traces

have been shown to be replayed in regions such as the hippocampus, medial prefrontal cortex, and visual cortex (Euston et al., 2007; Ji & Wilson, 2007; Skaggs & McNaughton, 1996). Potentially, this coordinated reactivation of memory traces is mediated by synchronous brain activity across brain regions, manifested in three tightly coupled cardinal brain oscillations, namely ripples (found primarily in the hippocampus), spindles (primarily found in the thalamus), and slow waves (that propagate from the prefrontal cortex to the hippocampus and medial temporal lobes) (Mölle et al., 2002; Nir et al., 2011; Sirota et al., 2003; Staresina et al., 2015). The visual cortex thus displays greater network participation during cueing, which could be due to greater connectivity with those regions that are involved in memory reprocessing. Indeed, we find increasing connectivity of the occipital cortex with the hippocampus, thalamus, and prefrontal cortex. This potentially resulted in a better embedding of the memory trace in the neocortex (we found some evidence for this in the increased recruitment of the parahippocampal gyrus in the post-sleep test). These results are consistent with cued memory reactivation mirroring endogenous memory reprocessing.

These results suggest that graph theory can be leveraged to detect subtle connectivity changes in response to memory reactivation that may capture increased interactions across hippocampal-neocortical networks. Accounting for whole-brain connectivity dynamics, as we did here, provides a unique part of the picture of memory reprocessing during sleep. Some open questions remain. As suggested above, the connectivity changes reported here may look a lot like endogenous memory reprocessing. However, it is important to know whether the connectivity profiles found here are specifically related to the cues. It could be that such memory reactivation is not specific at all, and that all memory traces associated with the same context are reactivated. This interpretation is certainly bolstered by the finding that the relation between memory stabilization and network participation in the early visual cortex was not specific to those object-locations associated with the presented cue sounds, but rather extended to all object-location associations found on the same grid. To assess specificity, future studies should look more specifically into patterns of brain activity during encoding, cued reactivation, and retrieval. As such, the specificity of cued reactivation could be assessed (for instance, using multivariate pattern analysis

methods to look at brain activity patterns measured with fMRI). Furthermore, it is of interest to know whether brain connectivity-changes in response to cueing are specific to certain electrophysiological brain states. This would allow us to better link the reported connectivity profile to brain physiology during slow-wave sleep. It might be possible to study if the connectivity changes reported here are specific to certain phases of oscillatory slow-waves, the so-called cortical 'up-states', where thalamic spindles and hippocampal ripples are facilitated and more likely to occur.

Memory reprocessing during sleep thus might serve to embed recently encoded memories gradually in the neocortical memory store, where it may be stabilized and protected from the catastrophic interference of new incoming information (McClelland, 2013). Along with the rapid integration of schema-consistent information, it might serve as an additional mechanism to train the neocortex on the most relevant information from awake experiences that were encoded initially through hippocampal-mediated mechanisms. Meanwhile, the bulk of this information is eventually forgotten (vivid details of an event, spatiotemporal context). Indeed, sleep supports memory retention in a selective manner, and information is prioritized based on perceived future relevance at encoding (van Dongen, Thielen, et al., 2012; Wilhelm et al., 2011), emotional salience (Hu et al., 2006; Payne et al., 2007; Wagner et al., 2006), consistency across episodes and with prior knowledge (Durrant et al., 2015; Tamminen et al., 2013). As such, memory reprocessing during slow-wave sleep might serve as an adaptive mechanism to selectively integrate the most relevant encoded information into a more permanent neocortical memory store.

## On the modulation of memory integration in the human brain

We can now turn to considering how the series of experimental results reported here contribute to answering our central research question, namely: ***How does the human brain selectively integrate information from incoming experience into the long-term memory store?*** There are several factors that modulate how incoming information is eventually integrated into the long-term memory store, at initial encoding (see **Chapter 2-4**), and during post-encoding sleep (see **Chapter 5**). These factors might influence the manner in which memories are represented in the brain in two different manners. In one instance, a memory trace of a learning

event is represented in a network of posterior association cortices representing low-level perceptual features, which are then bound together by the hippocampus into a coherent memory trace of an event including peripheral features, such as the spatial and temporal context (a hippocampal-neocortical network; see left panel of Figure 6.1). Alternatively, a memory trace of a learning event is primarily laid down in the neocortex (a neocortical network, see right panel of Figure 6.1). The work presented in this thesis builds on prominent theories that invoke hippocampal-neocortical versus neocortical networks, such as the standard model of memory consolidation (Squire, 1992; Squire & Zola-Morgan, 1991), or the complementary learning systems model (McClelland et al., 1995). A central premise of the studies in this thesis is, however, that there are several additional factors, beyond the mere passage of time and/or speed of learning, that may influence how incoming information is represented in these two types of networks, or any mix of them. At initial encoding, several such factors influence the extent to which a learning event is represented in a hippocampal-neocortical network, or a neocortical network. For instance, only a central feature or object in an emotional event may be encoded in a neocortical network with reduced involvement of the hippocampus, leading to the discarding of associated information. Alternatively, an event may be highly congruent with existing knowledge, in which case existing neocortical-based knowledge is recruited to embed certain features that are relevant, while peripheral information is discarded. After encoding, additional factors influence how information that may be initially dependent on a hippocampal-neocortical network may gradually become represented more independently of the hippocampus in a neocortical network. For instance, several events may be initially encoded in a hippocampal-neocortical network, but as regularities across events are detected, these regular features may become embedded in a neocortical network. Alternatively, events that are initially encoded in a hippocampal-neocortical network may become embedded in a neocortical network due to the selective, repeated reprocessing or replay of the underlying memory which strengthens neocortical connections, such as is the case during slow-wave sleep.

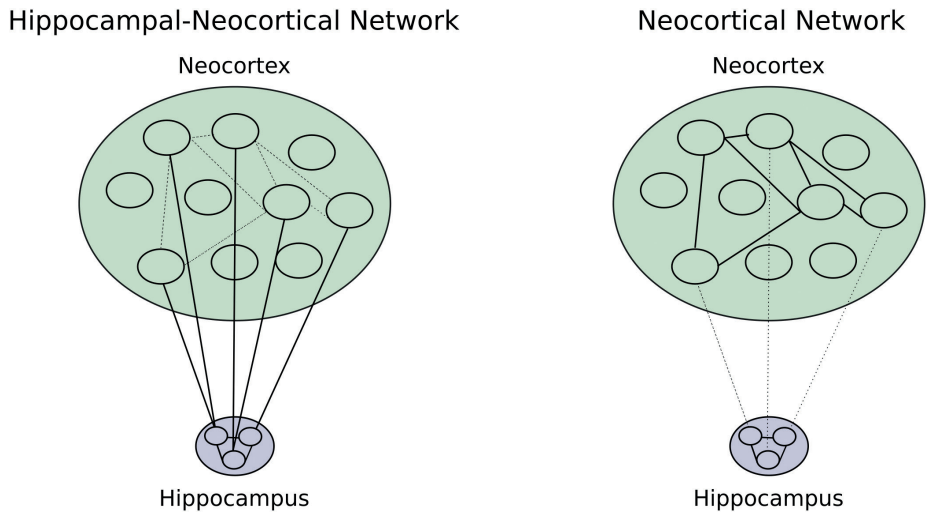



**Figure 6.1. Schematic display of two memory networks in the brain.**
*The left panel is a schematic depiction of a hippocampal-neocortical network, where various features of a memory trace that are represented in neocortical areas are bound together by the hippocampus. The right panel displays a neocortical network, where features represented in neocortical areas are well inter-connected and there are only weak connections to the hippocampus.*

In the context of this simplified model of two memory systems in the brain, I will now discuss how various factors that were investigated in this thesis may modulate the manner in which memories are integrated into the brain. It should be noted that this discussion is speculative, and provides a working thesis of how selective memory integration might work in the brain.

First of all, when encountering information with a negative valence (as discussed in Chapter 2), the amygdala may signal the valence to the mPFC, hippocampus, or both, promoting item encoding directly or through interconnections with neocortical regions in the medial temporal lobe (parahippocampal gyrus, see Ritchey et al., 2008). The mPFC is subsequently activated and downregulates encoding-related activity in the hippocampus. As such, only features related to the central gist are encoded in the neocortex, whereas associated peripheral information is not assimilated (see Figure 6.2A). This results in a potential emotional enhancement

of item and recognition memory, and an emotional impairment effect of associative memory and memory for peripheral contextual information.

A second factor concerns the congruency of incoming information with prior knowledge (as discussed in Chapter 3). If incoming information is congruent with a prior or activated schema, then this information is easily assimilated into the neocortical store. In such a case, hippocampal activity is downregulated and only certain features of individual events are encoded without their rich spatiotemporal context. Potentially, this process is set in motion by the medial prefrontal cortex, which detects resonating and mutually reinforcing activity in the neocortical network. Such schema instantiation in the neocortical networks sets in motion cortical plasticity, allowing information to be readily integrated but also for false features to be mutually reinforced and inadvertently encoded as part of the ongoing experience. When the medial prefrontal cortex is downregulated by TMS (as is presumably occurring in the reported study in Chapter 3), such heightened cortical plasticity in the neocortical network is mostly absent (see Figure 6.2B). As such, incoming information is less readily integrated into existing neocortical knowledge representations. Moreover, less memory errors occur, as new information is incorporated to a lesser extent in the context of pre-activated plastic neocortical networks that may inadvertently activate faulty information. Instead, the hippocampus is increasingly recruited to lay down encoded memory traces without explicitly linking them to prior knowledge.



A third factor concerns the situation when initially there is very little prior knowledge available that can serve as a neocortical scaffold to readily incorporate new information. Rather, across initial exposures, the hippocampus might be required to lay down associative information. Then, regularities across single exposures emerge and are detected (as discussed in Chapter 4), potentially through recurrent similarity computation and bidirectional interactions between the hippocampus and neocortical regions such as the medial prefrontal cortex (Kumaran & McClelland, 2012). These regularities accumulate in a neocortical region that may be suited to store such regularities. In the case of an associative knowledge structure comprising syntactical and semantic information (Hagoort, 2005), or more generally, complex sequences that conform to a finite state

grammar (Peigneux et al., 1999), this specialized region may be the left inferior frontal gyrus. Furthermore, the caudate nucleus might constitute a learning region that helps to update the neocortical associative memory store, and aids in the retrieval of recently updated knowledge that may not be as well-established yet in the neocortical store (Scimeca & Badre, 2012) (see Figure 6.2C).

The fourth factor that influences the manner in which knowledge is integrated into the long-term memory store plays a role in the offline period following initial learning. Presumably, memory traces are reprocessed during those offline periods, particularly during slow-wave sleep (Ji & Wilson, 2007; Skaggs & McNaughton, 1996), but also during wakeful rest (Carr et al., 2011). Visuo-spatial memories (such as discussed in Chapter 5) are presumably initially encoded in hippocampal-neocortical networks. Then during periods of slow-wave sleep, the memory traces that were initially laid down during encoding are repeatedly replayed during periods of synchronous firing across multiple brain regions, presumably including the hippocampus, thalamus, medial prefrontal cortex and posterior neocortical association cortices. The replaying of events might be occurring repeatedly across a period of time, and at a faster rate than the original learning experience unfolded (Euston et al., 2007). The rate of replay may be related to the associated reward (Ambrose, Pfeiffer, & Foster, 2016), perceived future relevance (van Dongen, Thielen, et al., 2012; Wilhelm et al., 2011), congruency with prior knowledge (Creery et al., 2015), or emotional arousal (Hu et al., 2006; Payne et al., 2007; Wagner et al., 2006). Moreover, replay-like reactivation of memories may be induced externally through 'Targeted Memory Reactivation' (TMR) using sounds or smells (Rasch et al., 2007; Rudoy et al., 2009). Such replay of these memory traces would strengthen connections amongst its nodes, including amongst the neocortical nodes themselves. After repeated replaying of certain memory traces, they then might become sufficiently embedded into the neocortical network, and the hippocampus is not needed anymore to bind relevant features of the memory trace (see Figure 6.2B).

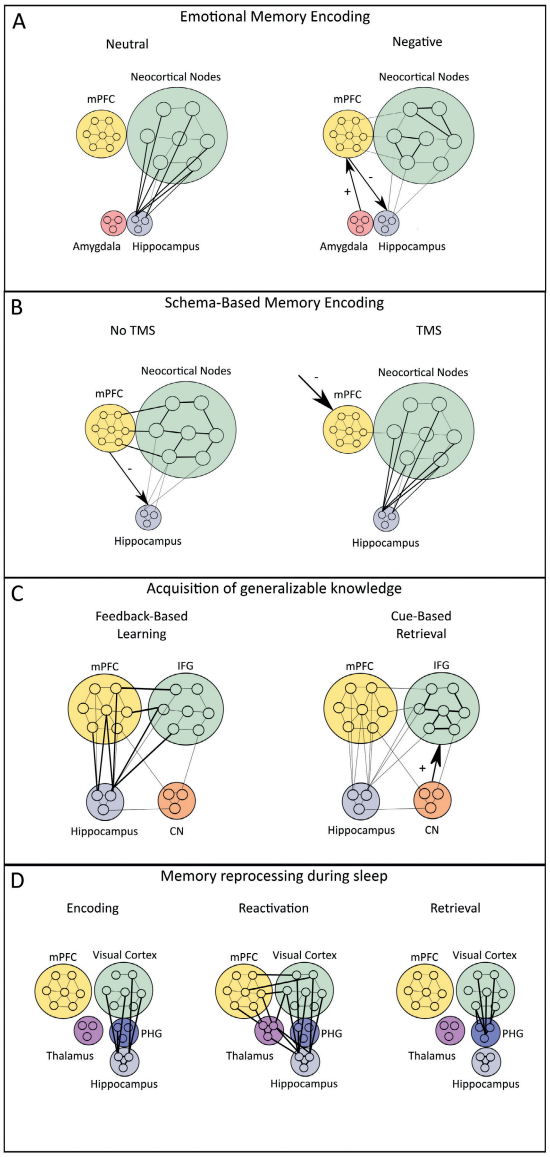
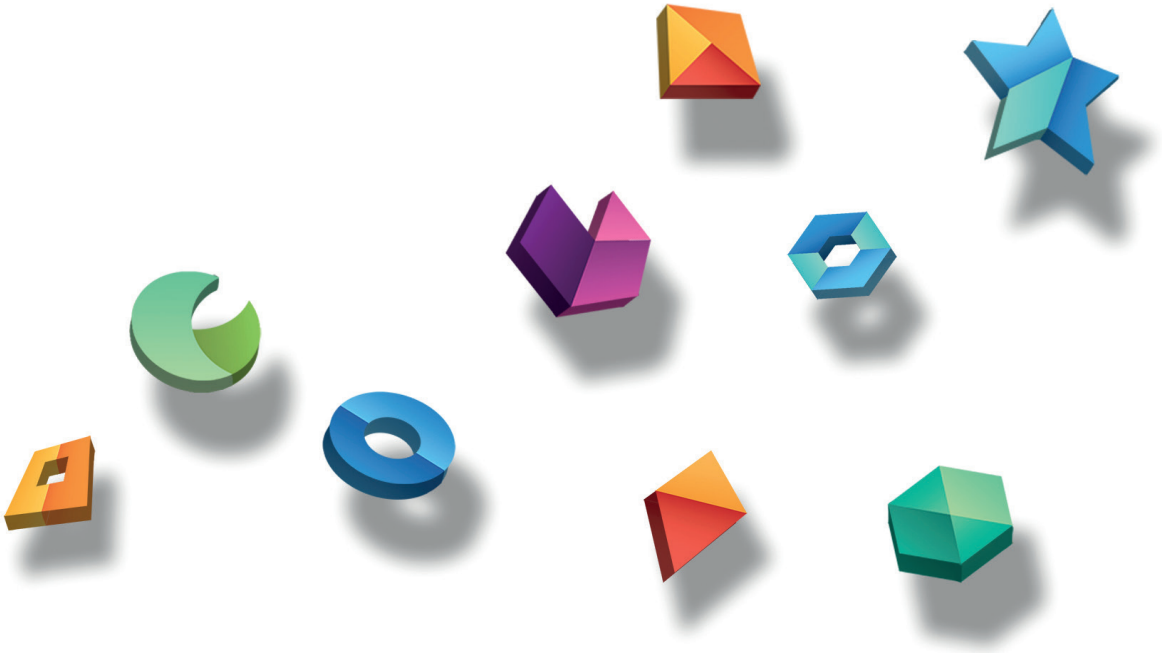



**Figure 6.2. Schematic model of modulatory influences on memory integration in the brain.**

*Small circles constitute neural modules within a brain area, whereas larger circles denote the entire brain area. Thin dotted lines constitute connections, whereas the thick lines constitute connections that are particularly active in a particular scenario. The scenarios depicted in panels A-D are explained in detail in the main text, and are speculative models of modulatory influences on memory integration in the human brain. Abbreviations: CN = Caudate Nucleus, IFG = Inferior Frontal Gyrus, mPFC = medial Prefrontal Cortex, PHG = parahippocampal gyrus.*

## Conclusion

The essential function of the declarative memory system is to make an informed simulation of the future, and to adapt decisions and behavior based on those simulations. The function of such memory storage is thus not primarily to accurately reflect the past, although that certainly helps, rather the purpose is to make a prediction of the future, whether this prediction ultimately proves to be accurate or not, and optimize behavior accordingly so as to promote survival and avoid threats and potentially stressful situations. For this purpose, our brains cannot, and do not need to store all the perceptual input they receive during every waking moment. Our brains would be overwhelmed and we would not be able to see the forest for the trees. Instead, our memory system developed mechanisms, through billions of years of evolution, to select information it deems relevant to maintain for future instances. As such, information is not simply copied verbatim into a long-term memory store, but heuristics are used to modulate how encoded experiences are selectively integrated into the existing memory store. These encoding heuristics are based on the emotional valence of incoming information, its congruency with an established schema, and regularities across episodes. Moreover, following initial encoding, the brain uses these same heuristics to selectively reprocess memories during sleep, providing another window for the brain to selectively integrate information into a long-term memory store. The studies reported in this thesis shed some light into how our brains achieve this impressive feat.



# CHAPTER

# A

## Appendix

References
Supplemental materials and methods
Research data management
Summary of thesis (EN & NL)
List of publications
Academic portfolio DGCN
Curriculum Vitae
Acknowledgements
Donders Graduate School for Cognitive Neuroscience

## References

A Adolphs, R., Tranel, D., & Buchanan, T. W. (2005). Amygdala damage impairs emotional memory for gist but not details of complex stimuli. *Nat Neurosci*, *8*(4), 512–518.

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioral and Brain Sciences*, *22*(3), 425–489.

Aggleton, J. P., Desimone, R., & Mishkin, M. (1986). The origin, course, and termination of the hippocampothalamic projections in the macaque. *Journal of Comparative Neurology*, *243*(3), 409–421.

Aguirre, G. K., Detre, J. A., Alsop, D. C., & D'Esposito, M. (1996). The parahippocampus subserves topographical learning in man. *Cerebral Cortex*, *6*(6), 823–829.

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*(2), 203.

Aleman-Gomez, Y., Melie-Garcia, L., & Valdes-Hernandez, P. A. (2006). IBASPM: toolbox for automatic parcellation of brain structures. In *12th Annual Meeting of the Organization for Human Brain Mapping. June 11-15, 2006. Florence, Italy*.

Amaral, D. G., & Cowan, W. M. (1980). Subcortical afferents to the hippocampal formation in the monkey. *Journal of Comparative Neurology*, *189*(4), 573–591.

Ambrose, R. E., Pfeiffer, B. E., & Foster, D. J. (2016). Reverse replay of hippocampal place cells is uniquely modulated by changing reward. *Neuron*, *91*(5), 1124–1136.

Aminoff, E., Gronau, N., & Bar, M. (2007). The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral Cortex*, *17*(7), 1493–1503.

Anderson, J. R. (1981). Effects of prior knowledge on memory for new information. *Memory & Cognition*, *9*(3), 237–246.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*(3), 409.

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron*, *65*(4), 550–562.

Arppe, A., Hendrix, P., Milin, P., Baayen, R. H., Sering, T., & Shaoul, C. (2015). ndl: Naive Discriminative Learning. R package version 0.2.17. Retrieved from https://cran.r-project.org/web/packages/ndl/

B Baayen, R. H., Piepenbrock, R., & Rijn, van H. (1993). The {CELEX} lexical data base on {CD-ROM}. Philadelphia, PA: Linguistic Data Consortium.

Backus, A. R., Bosch, S. E., Ekman, M., Grabovetsky, A. V., & Doeller, C. F. (2016). Mnemonic convergence in the human hippocampus. *Nature Communications*, *7*, 11991.

Backus, A. R., Schoffelen, J.-M., Szebényi, S., Hanslmayr, S., & Doeller, C. F. (2016). Hippocampal-prefrontal theta oscillations support memory integration. *Current Biology*, *26*(4), 450–457.

Badre, D., & Wagner, A. D. (2002). Semantic Retrieval, Mnemonic Control, and Prefrontal Cortex. *Behavioral and Cognitive Neuroscience Reviews* , *1*(3), 206–218.

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. L. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*, *319*(5870), 1640–1642.

Bandettini, P. A., & Ungerleider, L. G. (2001). From neuron to BOLD: new connections. *Nature Neuroscience*, *4*(9), 864.

Banks, S. J., Eddy, K. T., Angstadt, M., Nathan, P. J., & Phan, K. L. (2007). Amygdala–frontal connectivity during emotion regulation. *Social Cognitive and Affective Neuroscience* , *2*(4), 303–312.

Barbey, A. K., Krueger, F., & Grafman, J. (2009). Structured event complexes in the medial prefrontal cortex support counterfactual representations for future planning. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1521), 1291–1300.

Barker, A. T., Jalinous, R., & Freeston, I. L. (1985). Non-invasive magnetic stimulation of human motor cortex. *The Lancet*, *325*(8437), 1106–1107.

Barker, G. R. I., & Warburton, E. C. (2018). A critical role for the nucleus reuniens in long-term, but not short-term associative recognition memory formation. *Journal of Neuroscience*, 1802–1817.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1521), 1281–1289.

Bartlett, F. C. (1932). *Remembering: An experimental and social study. Cambridge: Cambridge University*.

Bassett, D. S., & Bullmore, E. D. (2006). Small-world brain networks. *The Neuroscientist*, *12*(6), 512–523.

Bassett, D. S., Bullmore, E. T., Meyer-Lindenberg, A., Apud, J. A., Weinberger, D. R., & Coppola, R. (2009). Cognitive fitness of cost-efficient brain functional networks. *Proceedings of the National Academy of Sciences*, *106*(28), 11747–11752.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295.

Bein, O., Reggev, N., & Maril, A. (2014). Prior knowledge influences on hippocampus and medial prefrontal cortex interactions in subsequent memory. *Neuropsychologia*, *64*, 320–330. https://doi.org/http://dx.doi.org/10.1016/j.neuropsychologia.2014.09.046

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Benoit, R. G. R. G., Gilbert, S. J. S. J., & Burgess, P. W. P. W. (2011). A Neural Mechanism Mediating the Impact of Episodic Prospection on Farsighted Decisions. *The Journal of Neuroscience*, *31*(18), 6771–6779.

Benoit, R. G., Szpunar, K. K., & Schacter, D. L. (2014). Ventromedial prefrontal cortex supports affective future simulation by integrating distributed knowledge. *Proceedings of the National Academy of Sciences*, *111*(46), 16550–16555.

Bergmann, T. O., Mölle, M., Diedrichs, J., Born, J., & Siebner, H. R. (2012). Sleep spindle-related reactivation of category-specific cortical regions after learning face-scene associations. *Neuroimage*, *59*(3), 2733–2742.

Berkers, R. M. W. J., Klumpers, F., & Fernández, G. (2016). Medial prefrontal–hippocampal connectivity during emotional memory encoding predicts individual differences in the loss of associative memory specificity. *Neurobiology of Learning and Memory*, *134*, 44–54.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex (New York, N.Y. : 1991)*, *19*(12), 2767–2796.

Bisby, J. A., & Burgess, N. (2014). Negative affect impairs associative memory but not item memory. *Learning & Memory* , *21*(1), 21–27.

Bisby, J. A., Horner, A. J., Hørlyck, L. D., & Burgess, N. (2016). Opposing effects of negative emotion on amygdalar and hippocampal memory for items and associations. *Social Cognitive and Affective Neuroscience*, *11*(6), 981–990.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine*, *34*(4), 537–541.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008.

Bluck, S., & Li, K. Z. H. (2001). Predicting memory completeness and accuracy: emotion and exposure in repeated autobiographical recall. *Applied Cognitive Psychology*, *15*(2), 145–158.

Boersma, M., Kemner, C., de Reus, M. A., Collin, G., Snijders, T. M., Hofman, D., … van den Heuvel, M. P. (2013). Disrupted functional brain networks in autistic toddlers. *Brain Connectivity*, *3*(1), 41–49.

Bonasia, K., Sekeres, M. J., Gilboa, A., Grady, C. L., Winocur, G., & Moscovitch, M. (2018). Prior knowledge modulates the neural substrates of encoding and retrieving naturalistic events at short and long delays. *Neurobiology of Learning and Memory*. 153A: 26-39.

Bonner, M. F., Peelle, J. E., Cook, P. A., & Grossman, M. (2013). Heteromodal conceptual processing in the angular gyrus. *NeuroImage*, *71*, 175–186.

Bonnì, S., Veniero, D., Mastropasqua, C., Ponzo, V., Caltagirone, C., Bozzali, M., & Koch, G. (2015). TMS evidence for a selective role of the precuneus in source memory retrieval. *Behavioural Brain Research*, *282*, 70–75.

Bonnici, H. M., Richter, F. R., Yazar, Y., & Simons, J. S. (2016). Multimodal Feature Integration in the Angular Gyrus during Episodic and Semantic Retrieval. *The Journal of Neuroscience* , *36*(20), 5462–5471.

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, *25*(1), 151–188.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, *8*(3), 323–343.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience*, *16*(13), 4207–4221.

Bradley, M. M., Greenwald, M. K., Petry, M. C., & Lang, P. J. (1992). Remembering pictures: pleasure and arousal in memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18*(2), 379–390.

Brady, T. F., & Oliva, A. (2008). Statistical Learning Using Real-World Scenes: Extracting Categorical Regularities Without Conscious Intent. *Psychological Science*, *19*(7), 678–685.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 717–726.

Brewin, C. R. (2007). Autobiographical memory for trauma: Update on four controversies. *Memory*, *15*(3), 227–248.

Brewin, C. R., Gregory, J. D., Lipton, M., & Burgess, N. (2010). Intrusive images in psychological disorders: characteristics, neural mechanisms, and treatment implications. *Psychological Review*, *117*(1), 210–232.

Brewin, C. R., Huntley, Z., & Whalley, M. G. (2012). Source memory errors associated with reports of posttraumatic flashbacks: A proof of concept study. *Cognition*, *124*(2), 234–238.

Brod, G., Lindenberger, U., Wagner, A. D., & Shing, Y. L. (2016). Knowledge Acquisition during Exam Preparation Improves Memory and Modulates Memory Formation. *The Journal of Neuroscience*, *36*(31), 8103 LP-8111.

Brod, G., Lindenberger, U., Werkle-Bergner, M., & Shing, Y. L. (2015). Differences in the neural signature of remembering schema-congruent and schema-incongruent events. *Neuroimage*, *117*, 358–366.

Brown, A. D., Root, J. C., Romano, T. A., Chang, L. J., Bryant, R. A., & Hirst, W. (2013). Overgeneralized autobiographical memory and future thinking in combat veterans with posttraumatic stress disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *44*(1), 129–134.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, *5*(1), 73–99.

Brown, R. W., Haacke, E. M., Cheng, Y.-C. N., Thompson, M. R., & Venkatesan, R. (2014). *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198.

Burgess, P. W., Scott, S. K., & Frith, C. D. (2003). The role of the rostral frontal cortex (area 10) in prospective memory: a lateral versus medial dissociation. *Neuropsychologia*, *41*(8), 906–918.

Burke, A., Heuer, F., & Reisberg, D. (1992). Remembering emotional events. *Memory & Cognition*, *20*(3), 277–290.

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, *315*(5820), 1860–1862.

Buzsáki, G. (1996). The hippocampo-neocortical dialogue. *Cerebral Cortex*, *6*(2), 81–92.

C

Cabeza, R., Rao, S. M., Wagner, A. D., Mayer, A. R., & Schacter, D. L. (2001). Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study of veridical and illusory recognition memory. *Proceedings of the National Academy of Sciences* , *98*(8), 4805–4810.

Cahill, L., Babinsky, R., Markowitsch, H. J., & McGaugh, J. L. (1995). The amygdala and emotional memory. *Nature*, *377*(6547), 295–296.

Cahill, L., Haier, R. J., Fallon, J., Alkire, M. T., Tang, C., Keator, D., … McGaugh, J. L. (1996). Amygdala activity at encoding correlated with long-term, free recall of emotional information. *Proceedings of the National Academy of Sciences*, *93*(15), 8016–8021.

Cairney, S. A., Durrant, S. J., Hulleman, J., & Lewis, P. A. (2014). Targeted memory reactivation during slow wave sleep facilitates emotional memory consolidation. *Sleep*, *37*(4), 701–707.

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D., & Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *The Journal of Neuroscience*, *20*(19), RC99.

Cao, H., Plichta, M. M., Schäfer, A., Haddad, L., Grimm, O., Schneider, M., … Tost, H. (2014). Test–retest reliability of fMRI-based graph theoretical properties during working memory, emotion processing, and resting state. *Neuroimage*, *84*, 888–900.

Carr, M. F., Jadhav, S. P., & Frank, L. M. (2011). Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature Neuroscience*, *14*(2), 147.

Carr, M. F., Karlsson, M. P., & Frank, L. M. (2012). Transient slow gamma synchrony underlies hippocampal memory replay. *Neuron*, *75*(4), 700–713.

Cassel, J.-C., De Vasconcelos, A. P., Loureiro, M., Cholvin, T., Dalrymple-Alford, J. C., & Vertes, R. P. (2013). The reuniens and rhomboid nuclei: neuroanatomy, electrophysiological characteristics and behavioral implications. *Progress in Neurobiology*, *111*, 34–52.

Chessa, A. G., & Murre, J. M. J. (2007). A Neurocognitive Model of Advertisement Content and Brand Name Recall. *Marketing Science*, *26*(1), 130–141.

Christianson, S.-Å., & Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Cognition and Emotion*, *5*(2), 81–108.

Christianson, S. A. (1992). Emotional stress and eyewitness memory: a critical review. *Psychological Bulletin*, *112*(2), 284–309.

Collin, S. H. P., Milivojevic, B., & Doeller, C. F. (2015). Memory hierarchies map onto the hippocampal long axis in humans. *Nature Neuroscience*, *18*(11), 1562–1566.

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, *352*(6292), 1464–1468.

Corkin, S. (1968). Acquisition of motor skill after bilateral medial temporal-lobe excision. *Neuropsychologia*, *6*(3), 255–265.

Corkin, S. (1984). Lasting consequences of bilateral medial temporal lobectomy: Clinical course and experimental findings in HM. In *Seminars in Neurology* (Vol. 4, pp. 249–259).

Corkin, S. (2002). What's new with the amnesic patient HM? *Nature Reviews Neuroscience*, *3*(2), 153.

Creery, J. D., Oudiette, D., Antony, J. W., & Paller, K. A. (2015). Targeted memory reactivation during sleep depends on prior learning. *Sleep*, *38*(5), 755–763.

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, *7*(9), 415–423.

D

Dalgleish, T., Williams, J. M. G., Golden, A.-M. J., Perkins, N., Barrett, L. F., Barnard, P. J., … Watkins, E. (2007). Reduced specificity of autobiographical memory and depression: the role of executive control. *Journal of Experimental Psychology. General*, *136*(1), 23–42.

Dalton, M. A., Weickert, T. W., Hodges, J. R., Piguet, O., & Hornberger, M. (2013). Impaired acquisition rates of probabilistic associative learning in frontotemporal dementia is associated with fronto-striatal atrophy. *NeuroImage: Clinical*, *2*, 56–62.

Danker, J. F., & Anderson, J. R. (2010). The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin*, *136*(1), 87.

Danks, D. (2003). Equilibria of the Rescorla–Wagner model. *Journal of Mathematical Psychology*, *47*(2), 109–121.

Darsaud, A., Dehon, H., Lahl, O., Sterpenich, V., Boly, M., Dang-Vu, T., … Peters, F. (2011). Does sleep promote false memories? *Journal of Cognitive Neuroscience*, *23*(1), 26–40.

Davidson, T. J., Kloosterman, F., & Wilson, M. A. (2009). Hippocampal replay of extended experience. *Neuron*, *63*(4), 497–507.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci*, *8*(12), 1704–1711.

de Voogd, L. D., Klumpers, F., Fernández, G., & Hermans, E. J. (2017). Intrinsic functional connectivity between amygdala and hippocampus during rest predicts enhanced memory under stress. *Psychoneuroendocrinology*, *75*, 192–202.

Deco, G., Jirsa, V. K., & McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nature Reviews Neuroscience*, *12*(1), 43.

Deng, Z.-D., Lisanby, S. H., & Peterchev, A. V. (2013). Electric field depth–focality tradeoff in transcranial magnetic stimulation: simulation comparison of 50 coil designs. *Brain Stimulation*, *6*(1), 1–13.

Deng, Z.-D., Lisanby, S. H., & Peterchev, A. V. (2014). Coil design considerations for deep transcranial magnetic stimulation. *Clinical Neurophysiology*, *125*(6), 1202–12.

DeVito, J. L. (1980). Subcortical projections to the hippocampal formation in squirrel monkey (Saimiri sciureus). *Brain Research Bulletin*, *5*(3), 285–289.

DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J. O. N., Cox, R., … Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences*, *93*(6), 2382–2386.

Diba, K., & Buzsáki, G. (2007). Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience*, *10*(10), 1241–1242.

Diedrichsen, J., Yokoi, A., & Arbuckle, S. A. (2018). Pattern component modeling: A flexible approach for understanding the representational structure of brain activity patterns. *NeuroImage*, *180*, 119–133.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience*, *11*(2), 114–126.

Diekelmann, S., Born, J., & Wagner, U. (2010). Sleep enhances false memories depending on general memory performance. *Behavioural Brain Research*, *208*(2), 425–429.

Diekelmann, S., Büchel, C., Born, J., & Rasch, B. (2011). Labile or stable: opposing consequences for memory when reactivated during waking and sleep. *Nature Neuroscience*, *14*(3), 381–386.

Doeller, C. F., & Burgess, N. (2008). Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(15), 5909–5914.

Doeller, C. F., King, J. A., & Burgess, N. (2008). Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proceedings of the National Academy of Sciences*, *105*(15), 5915–5920.

Dolcos, F., LaBar, K. S., & Cabeza, R. (2004). Interaction between the Amygdala and the Medial Temporal Lobe Memory System Predicts Better Memory for Emotional Events. *Neuron*, *42*(5), 855–863.

Dolleman-Van der Weel, M. J., Lopes da Silva, F. H., & Witter, M. P. (1997). Nucleus Reuniens Thalami Modulates Activity in Hippocampal Field CA1 through Excitatory and Inhibitory Mechanisms. *The Journal of Neuroscience*, *17*(14), 5640–5650.

Dominique, J.-F., & Papassotiropoulos, A. (2006). Identification of a genetic cluster influencing memory performance and hippocampal activity in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(11), 4270–4274.

Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., … Raichle, M. E. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proceedings of the National Academy of Sciences*, *104*(26), 11073–11078.

Dupret, D., O'Neill, J., Pleydell-Bouverie, B., & Csicsvari, J. (2010). The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nature Neuroscience*, *13*(8), 995–1002.

Durrant, S. J., Cairney, S. A., McDermott, C., & Lewis, P. A. (2015). Schema-conformant memories are preferentially consolidated during REM sleep. *Neurobiology of Learning and Memory*, *122*, 41–50.

Durrant, S. J., Taylor, C., Cairney, S., & Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, *49*(5), 1322–1331.

E

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, *66*(3), 183.

Ehlers, A., Hackmann, A., & Michael, T. (2004). Intrusive re-experiencing in post-traumatic stress disorder: Phenomenology, theory, and therapy. *Memory*, *12*(4), 403–415.

Eichenbaum, H. (2000). A cortical-hippocampal system for declarative memory. *Nat Rev Neurosci*, *1*(1), 41–50.

Eichenbaum, H., Otto, T., & Cohen, N. J. (1994). Two functional components of the hippocampal memory system. *Behavioral and Brain Sciences*, *17*(3), 449–472.

Ekman, M., Derrfuss, J., Tittgemeyer, M., & Fiebach, C. J. (2012). Predicting errors from reconfiguration patterns in human brain networks. *Proceedings of the National Academy of Sciences*, *109*(41), 16714–16719.

Ekman, M., Kok, P., & de Lange, F. P. (2017). Time-compressed preplay of anticipated events in human primary visual cortex. *Nature Communications*, *8*.

Eldaief, M. C., Halko, M. a, Buckner, R. L., & Pascual-Leone, A. (2011). Transcranial magnetic stimulation modulates the brain's intrinsic activity in a frequency-dependent manner. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(52), 21229–34.

Epstein, C. M., Sekino, M., Yamaguchi, K., Kamiya, S., & Ueno, S. (2002). Asymmetries of prefrontal cortex in human episodic memory: effects of transcranial magnetic stimulation on learning abstract patterns. *Neuroscience Letters*, *320*(1), 5–8.

Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, *23*(1), 115–125.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Erk, S., Kiefer, M., Grothe, J., Wunderlich, A. P., Spitzer, M., & Walter, H. (2003). Emotional context modulates subsequent memory effect. *NeuroImage*, *18*(2), 439–447.

Erk, S., Meyer-Lindenberg, A., Schnell, K., von Boberfeld, C. O., Esslinger, C., Kirsch, P., … Witt, S. H. (2010). Brain function in carriers of a genome-wide supported bipolar disorder variant. *Archives of General Psychiatry*, *67*(8), 803–811.

Erk, S., Spottke, A., Meisen, A., Wagner, M., Walter, H., & Jessen, F. (2011). Evidence of neuronal compensation during episodic memory in subjective memory impairment. *Archives of General Psychiatry*, *68*(8), 845–852.

Etkin, A., Egner, T., & Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Sciences*, *15*(2), 85–93.

Euston, D. R., Tatsuno, M., & McNaughton, B. L. (2007). Fast-forward playback of recent memory sequences in prefrontal cortex during sleep. *Science*, *318*(5853), 1147–1150.

F Farovik, A., Place, R. J., Miller, D. R., & Eichenbaum, H. (2011). Amygdala lesions selectively impair familiarity in recognition memory. *Nature Neuroscience*, *14*(11), 1416–1417.

Fastenrath, M., Coynel, D., Spalek, K., Milnik, A., Gschwind, L., Roozendaal, B., … de Quervain, D. J. F. (2014). Dynamic modulation of amygdala–hippocampal connectivity by emotional arousal. *Journal of Neuroscience*, *34*(42), 13935–13947.

Finsterwald, C., & Alberini, C. M. (2014). Stress and glucocorticoid receptor-dependent mechanisms in long-term memory: from adaptive responses to psychopathologies. *Neurobiology of Learning and Memory*, *112*, 17–29.

Fitzgerald, P. B., Fountain, S., & Daskalakis, Z. J. (2006). A comprehensive review of the effects of rTMS on motor cortical excitability and inhibition. *Clinical Neurophysiology*, *117*(12), 2584–2596.

Floel, A., Poeppel, D., Buffalo, E. A., Braun, A., Wu, C. W.-H., Seo, H.-J., … Cohen, L. G. (2004). Prefrontal cortex asymmetry for memory encoding of words and abstract shapes. *Cerebral Cortex*, *14*(4), 404–409.

Foa, E. B., Gilboa-Schechtman, E., Amir, N., & Freshman, M. (2000). Memory bias in generalized social phobia: Remembering negative emotional expressions. *Journal of Anxiety Disorders*, *14*(5), 501–519.

Forkstam, C., Hagoort, P., Fernández, G., Ingvar, M., & Petersson, K. M. (2006). Neural correlates of artificial syntactic structure classification. *Neuroimage*, *32*(2), 956–967.

Fowler, M. J., Sullivan, M. J., & Ekstrand, B. R. (1973). Sleep and memory. *Science*, *179*(4070), 302–304.

Fox, M. D., Halko, M. A., Eldaief, M. C., & Pascual-Leone, A. (2012). Measuring and manipulating brain connectivity with resting state functional connectivity magnetic resonance imaging (fcMRI) and transcranial magnetic stimulation (TMS). *Neuroimage*, *62*(4), 2232–2243.

Frey, U., & Morris, R. G. M. (1997). Synaptic tagging and long-term potentiation. *Nature*, *385*(6616), 533–536.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*. US: American Psychological Association.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and Modulatory Interactions in Neuroimaging. *NeuroImage*, *6*(3), 218–229.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage*, *19*(4), 1273–1302.

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N. A., & Dale, A. M. (1999). Stochastic designs in event-related fMRI. *Neuroimage*, *10*(5), 607–619.

G

Geday, J., Gjedde, A., Boldsen, A.-S., & Kupers, R. (2003). Emotional valence modulates activity in the posterior fusiform gyrus and inferior medial prefrontal cortex in social perception. *NeuroImage*, *18*(3), 675–684.

Geday, J., Kupers, R., & Gjedde, A. (2007). As time goes by: temporal constraints on emotional activation of inferior medial prefrontal cortex. *Cerebral Cortex*, *17*(12), 2753–2759.

Gelfand, J. R., & Bookheimer, S. Y. (2003). Dissociating neural mechanisms of temporal sequencing and processing phonemes. *Neuron*, *38*(5), 831–842.

Genzel, L., & Battaglia, F. P. (2017). Cortico-Hippocampal Circuits for Memory Consolidation: The Role of the Prefrontal Cortex. In *Cognitive Neuroscience of Memory Consolidation* (pp. 265–281). Springer.

Genzel, L., Kroes, M. C. W., Dresler, M., & Battaglia, F. P. (2014). Light sleep versus slow wave sleep in memory consolidation: a question of global versus local processes? *Trends in Neurosciences*, *37*(1), 10–19.

Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, *53*, 104–14.

Gilboa, A., & Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. *Trends in Cognitive Sciences*, *21*(8), 618–631.

Girardeau, G., Benchenane, K., Wiener, S. I., Buzsáki, G., & Zugaro, M. B. (2009). Selective suppression of hippocampal ripples impairs spatial memory. *Nature Neuroscience*, *12*(10), 1222–1223.

Goodman, G. S. (1980). Picture memory: How the action schema affects retention. *Cognitive Psychology*, *12*(4), 473–495.

Grahn, J. A., Parkinson, J. A., & Owen, A. M. (2008). The cognitive functions of the caudate nucleus. *Progress in Neurobiology*, *86*(3), 141–155.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, *100*(1), 253–258.

Grover, V. P. B., Tognarelli, J. M., Crossey, M. M. E., Cox, I. J., Taylor-Robinson, S. D., & McPhail, M. J. W. (2015). Magnetic resonance imaging: principles and techniques: lessons for clinicians. *Journal of Clinical and Experimental Hepatology*, *5*(3), 246–255.

Guimera, R., Mossa, S., Turtschi, A., & Amaral, L. A. N. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences*, *102*(22), 7794–7799.

Guise, K. G., & Shapiro, M. L. (2017). Medial prefrontal cortex reduces memory interference by modifying hippocampal encoding. *Neuron*, *94*(1), 183–192.

Gusnard, D. A., Akbudak, E., Shulman, G. L., & Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(7), 4259–4264.

Gutchess, A. H., & Schacter, D. L. (2012). The neural correlates of gist-based true and false recognition. *NeuroImage*, *59*(4), 3418–3426.

H

Hagoort, P. (2003). How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage*, *20*, S18–S29.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, *9*(9), 416–423.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438–441.

Halai, A. D., Parkes, L. M., & Welbourne, S. R. (2015). Dual-echo fMRI can detect activations in inferior temporal lobe during intelligible speech comprehension. *NeuroImage*, *122*, 214–221.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences*, *5*(9), 394–400.

Hamann, S. B., Ely, T. D., Grafton, S. T., & Kilts, C. D. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nature Neuroscience*, *2*(3), 289–293.

Hanslmayr, S., Matuschek, J., & Fellner, M.-C. (2014). Entrainment of prefrontal beta oscillations induces an endogenous echo and impairs memory formation. *Current Biology*, *24*(8), 904–909.

Haruno, M., & Kawato, M. (2006). Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. *Journal of Neurophysiology*, *95*(2), 948–959.

Haruno, M., Kuroda, T., Doya, K., Toyama, K., Kimura, M., Samejima, K., … Kawato, M. (2004). A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task. *Journal of Neuroscience*, *24*(7), 1660–1665.

Hassabis, D., & Maguire, E. A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1263–1271.

Hayama, H. R., Vilberg, K. L., & Rugg, M. D. (2012). Overlap between the neural correlates of cued recall and source memory: evidence for a generic recollection network? *Journal of Cognitive Neuroscience*, *24*(5), 1127–1137.

Haynes, J.-D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, *8*(5), 686.

Hayward, G., Mehta, M. A., Harmer, C., Spinks, T. J., Grasby, P. M., & Goodwin, G. M. (2007). Exploring the physiological effects of double-cone coil TMS over the medial frontal cortex on the anterior cingulate cortex: an H215O PET study. *European Journal of Neuroscience*, *25*(7), 2224–2233.

Hennies, N., Ralph, M. A. L., Kempkes, M., Cousins, J. N., & Lewis, P. A. (2016). Sleep spindle density predicts the effect of prior knowledge on memory consolidation. *Journal of Neuroscience*, *36*(13), 3799–3810.

Henson, R. N., Campbell, K. L., Davis, S. W., Taylor, J. R., Emery, T., Erzinclioglu, S., … Calder, A. C. (2016). Multiple determinants of lifespan memory differences. *Scientific Reports*, *6*, 32527.

Hermans, E. J., Kanen, J. W., Tambini, A., Fernández, G., Davachi, L., & Phelps, E. A. (2017). Persistence of amygdala–hippocampal connectivity and multi-voxel correlation structures during awake rest after fear learning predicts long-term expression of fear. *Cerebral Cortex*, *27*(5), 3028–3041.

Holl, A. K., Wilkinson, L., Tabrizi, S. J., Painold, A., & Jahanshahi, M. (2012). Probabilistic classification learning with corrective feedback is selectively impaired in early Huntington's disease—Evidence for the role of the striatum in learning with feedback. *Neuropsychologia*, *50*(9), 2176–2186.

Hoover, W. B., & Vertes, R. P. (2012). Collateral projections from nucleus reuniens of thalamus to hippocampus and medial prefrontal cortex in the rat: a single and double retrograde fluorescent labeling study. *Brain Structure and Function*, *217*(2), 191–209.

Hope, L., & Wright, D. (2007). Beyond unusual? Examining the role of attention in the weapon focus effect. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *21*(7), 951–961.

Howell, D. (2012). *Statistical methods for psychology*. Cengage Learning.

Hsu, D. T., & Price, J. L. (2007). Midline and intralaminar thalamic connections with the orbital and medial prefrontal networks in macaque monkeys. *Journal of Comparative Neurology*, *504*(2), 89–111.

Hu, P., Stylos-Allan, M., & Walker, M. P. (2006). Sleep facilitates consolidation of emotional declarative memory. *Psychological Science*, *17*(10), 891–898.

Huang, Y.-Z., Edwards, M. J., Rounis, E., Bhatia, K. P., & Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron*, *45*(2), 201–206.

Huang, Y.-Z., Rothwell, J. C., Chen, R.-S., Lu, C.-S., & Chuang, W.-L. (2011). The theoretical model of theta burst form of repetitive transcranial magnetic stimulation. *Clinical Neurophysiology*, *122*(5),

Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging* (Vol. 1). Sinauer Associates Sunderland, MA.

Hyndman, R., Athanasopoulos, G., Razbash, S., Schmidt, D., Zhou, Z., & Khan, Y. (2013). forecast: Forecasting functions for time series and linear models [R package version 4.03].

I Iber, C., Ancoli-Israel, S., Chesson, A. L., & Quan, S. F. (2007). The new Sleep Scoring Manual–The evidence behind the rules. *Journal of Clinical Sleep Medicine*, *3*(2), 107.

J Jai, Y. Y., Liu, D. F., Loback, A., Grossrubatscher, I., & Frank, L. M. (2018). Specific hippocampal representations are linked to generalized cortical representations in memory. *Nature Communications*, *9*(1), 2209.

Javadi, A. H., Cheng, P., & Walsh, V. (2012). Short duration transcranial direct current stimulation (tDCS) modulates verbal memory. *Brain Stimulation*, *5*(4), 468–474.

Javadi, A. H., & Walsh, V. (2012). Transcranial direct current stimulation (tDCS) of the left dorsolateral prefrontal cortex modulates declarative memory. *Brain Stimulation*, *5*(3), 231–241.

Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature Neuroscience*, *10*(1), 100–107.

Jung, J., Bungert, A., Bowtell, R., & Jackson, S. R. (2016). Vertex stimulation as a control site for transcranial magnetic stimulation: A concurrent TMS/fMRI study. *Brain Stimulation*, *9*(1), 58–64.

K Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679.

Kemp, A., & Manahan-Vaughan, D. (2004). Hippocampal long-term depression and long-term potentiation encode different aspects of novelty acquisition. *Proceedings of the National Academy of Sciences*, *101*(21), 8192–8197.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007). Effects of emotion on memory specificity: Memory trade-offs elicited by negative visually arousing stimuli. *Journal of Memory and Language*, *56*(4), 575–591.

Kensinger, E. A., & Schacter, D. L. (2006). Amygdala activity is associated with the successful encoding of item, but not source, information for positive and negative stimuli. *Journal of Neuroscience*, *26*(9), 2564–2570.

Kensinger, E. A., Ullman, M. T., & Corkin, S. (2001). Bilateral medial temporal lobe damage does not affect lexical or grammatical processing: Evidence from amnesic patient HM. *Hippocampus*, *11*(4), 347–360.

Kheirbek, M. A., Klemenhagen, K. C., Sahay, A., & Hen, R. (2012). Neurogenesis and generalization: a new approach to stratify and treat anxiety disorders. *Nature Neuroscience*, *15*(12), 1613.

Kilpatrick, L., & Cahill, L. (2003). Amygdala modulation of parahippocampal and frontal regions during emotionally influenced memory storage. *NeuroImage*, *20*(4), 2091–2099.

Kim, H., & Cabeza, R. (2007). Differential contributions of prefrontal, medial temporal, and sensory-perceptual regions to true and false memory formation. *Cerebral Cortex*, *17*(9), 2143–2150.

King, M. J., MacDougall, A. G., Ferris, S. M., Levine, B., MacQueen, G. M., & McKinnon, M. C. (2010). A review of factors that moderate autobiographical memory performance in patients with major depressive disorder. *Journal of Clinical and Experimental Neuropsychology*, *32*(10), 1122–1144.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(31), 10681–10686.

Klucharev, V., Munneke, M. A. M., Smidts, A., & Fernández, G. (2011). Downregulation of the posterior medial frontal cortex prevents social conformity. *The Journal of Neuroscience*, *31*(33), 11934–40.

Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *1*(2), 106–120. Retrieved from

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in Cognitive Sciences*, *20*(7), 512–534.

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. *Psychological Review*, *119*(3), 573–616.

Kumaran, D., Melo, H. L., & Duzel, E. (2012). The emergence and representation of knowledge about social and non-social hierarchies. *Neuron*, *76*(3), 653–666.

Kumaran, D., Summerfield, J. J., Hassabis, D., & Maguire, E. A. (2009). Tracking the emergence of conceptual knowledge during human decision making. *Neuron*, *63*(6), 889–901.

L

Laakso, I., Hirata, A., & Ugawa, Y. (2013). Effects of coil orientation on the electric field induced by TMS over the hand motor area. *Physics in Medicine and Biology*, *59*(1), 203.

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews. Neuroscience*, *7*(1), 54–64.

Laney, C., Campbell, H. V, Heuer, F., & Reisberg, D. (2004). Memory for thematically arousing events. *Memory & Cognition*, *32*(7), 1149–1159.

Lansink, C. S., Goltstein, P. M., Lankelma, J. V, McNaughton, B. L., & Pennartz, C. M. A. (2009). Hippocampus leads ventral striatum in replay of place-reward information. *PLoS Biology*, *7*(8), e1000173.

Latchoumane, C.-F. V, Ngo, H.-V. V, Born, J., & Shin, H.-S. (2017). Thalamic spindles promote memory formation during sleep through triple phase-locking of cortical, thalamic, and hippocampal rhythms. *Neuron*, *95*(2), 424–435.

Lee, A. S., Duman, R. S., & Pittenger, C. (2008). A double dissociation revealing bidirectional competition between striatum and hippocampus during learning. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(44), 17163–17168.

Lemogne, C., Delaveau, P., Freton, M., Guionnet, S., & Fossati, P. (2012). Medial prefrontal cortex and the self in major depression. *Journal of Affective Disorders*, *136*(1–2), e1–e11.

Leutgeb, J. K., Leutgeb, S., Moser, M.-B., & Moser, E. I. (2007). Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science*, *315*(5814), 961–966.

Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M.-B., & Moser, E. I. (2004). Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science*, *305*(5688), 1295–1298.

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences*, *15*(8), 343–351.

Lin, D. (1998). An information-theoretic definition of similarity. In J. W. Shavlik (Ed.), *Proceedings of the Fifteenth International Conference on Machine Learning* (pp. 296–304). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., & Pine, D. S. (2005). Classical fear conditioning in the anxiety disorders: a meta-analysis. *Behaviour Research and Therapy*, *43*(11), 1391–1424.

Liu, H., Qin, W., Li, W., Fan, L., Wang, J., Jiang, T., & Yu, C. (2013). Connectivity-Based Parcellation of the Human Frontal Pole with Diffusion Tensor Imaging. *The Journal of Neuroscience*, *33*(16), 6782 LP-6790.

Liu, T. T., Frank, L. R., Wong, E. C., & Buxton, R. B. (2001). Detection power, estimation efficiency, and predictability in event-related fMRI. *Neuroimage*, *13*(4), 759–773.

Liu, Z.-X., Grady, C., & Moscovitch, M. (2017). Effects of Prior-Knowledge on Brain Activation and Connectivity During Associative Memory Encoding. *Cerebral Cortex, 27*(3), 1991-2009.

Loftus, E. F. (1996). Memory distortion and false memory creation. *Journal of the American Academy of Psychiatry and the Law Online*, *24*(3), 281–295.

Logothetis, N. K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *Journal of Neuroscience*, *23*(10), 3963–3971.

M Macmillan, N. A., & Creelman, C. D. (2005). Detection theory. *A User's Guide*.

Madan, C. R., Caplan, J. B., Lau, C. S. M., & Fujiwara, E. (2012). Emotional arousal does not enhance association-memory. *Journal of Memory and Language*, *66*(4), 695–716.

Maguire, E. A., & Mullally, S. L. (2013). The hippocampus: a manifesto for change. *Journal of Experimental Psychology: General*, *142*(4), 1180.

Mahan, A. L., & Ressler, K. J. (2012). Fear conditioning, synaptic plasticity and the amygdala: implications for posttraumatic stress disorder. *Trends in Neurosciences*, *35*(1), 24–35.

Markman, E. M., & Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive Psychology*, *16*(1), 1–27.

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences*, *11*(10), 442–450.

Massimini, M., Huber, R., Ferrarelli, F., Hill, S., & Tononi, G. (2004). The sleep slow oscillation as a traveling wave. *Journal of Neuroscience*, *24*(31), 6862–6870.

Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science*, *2*(1), 33–52.

Mather, M., Gorlick, M. A., & Nesmith, K. (2009). The limits of arousal's memory-impairing effects on nearby information. *The American Journal of Psychology*, *122*(3), 349–369. Retrieved from

Mather, M., & Knight, M. (2008). The emotional harbinger effect: poor context memory for cues that previously predicted something arousing. *Emotion (Washington, D.C.)*, *8*(6), 850–860.

McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition*, *30*(7), 1065–1077.

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology. General*, *142*(4), 1190–1210.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. *Science*, *306*(5695), 503–507.

McDermott, K. B., Gilmore, A. W., Nelson, S. M., Watson, J. M., & Ojemann, J. G. (2017). The parietal memory network activates similarly for true and associative false recognition elicited via the DRM procedure. *Cortex, 87,* 96-107.

McDonald, R. J., & White, N. M. (1993). A triple dissociation of memory systems: hippocampus, amygdala, and dorsal striatum. *Behavioral Neuroscience*, *107*(1), 3–22.

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience*, *27*(1), 1–28.

McKeon, S., Pace-Schott, E. F., & Spencer, R. M. C. (2012). Interaction of sleep and emotional content on the production of false memories. *PloS One*, *7*(11), e49353.

Melo, B., Winocur, G., & Moscovitch, M. (1999). False recall and false recognition: An examination of the effects of selective and combined lesions to the medial temporal lobe/diencephalon and frontal lobe structures. *Cognitive Neuropsychology*, *16*(3–5), 343–359.

Migueles, M., & Garcia-Bajos, E. (1999). Recall, recognition, and confidence patterns in eyewitness testimony. *Applied Cognitive Psychology*, *13*(3), 257–268.

Miller, E. K., Freedman, D. J., & Wallis, J. D. (2002). The prefrontal cortex: categories, concepts and cognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *357*(1424), 1123–1136.

Milner, B., Corkin, S., & Teuber, H.-L. (1968). Further analysis of the hippocampal amnesic syndrome: 14-year follow-up study of HM. *Neuropsychologia*, *6*(3), 215–234.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *17*(8), 1306–1315.

Mitchell, J. P., Neil Macrae, C., & Banaji, M. R. (2005). Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *NeuroImage*, *26*(1), 251–257.

Mitra, A., Snyder, A. Z., Hacker, C. D., Pahwa, M., Tagliazucchi, E., Laufs, H., ... Raichle, M. E. (2016). Human cortical–hippocampal dialogue in wake and slow-wave sleep. *Proceedings of the National Academy of Sciences, 113*(44), E6868-E6876.

Mölle, M., Marshall, L., Gais, S., & Born, J. (2002). Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep. *Journal of Neuroscience*, *22*(24), 10941–10947.

Mölle, M., Yeshenko, O., Marshall, L., Sara, S. J., & Born, J. (2006). Hippocampal sharp wave-ripples linked to slow oscillations in rat slow-wave sleep. *Journal of Neurophysiology*, *96*(1), 62–70.

Moradi, A. R., Taghavi, R., Neshat-Doost, H. T., Yule, W., & Dalgleish, T. (2000). Memory bias for emotional information in children and adolescents with posttraumatic stress disorder: A preliminary study. *Journal of Anxiety Disorders*, *14*(5), 521–534.

Morrissey, M. D., Insel, N., & Takehara-Nishiuchi, K. (2017). Generalizable knowledge outweighs incidental details in prefrontal ensemble code over time. *Elife*, *6*, e22177.

Moscovitch, M. (2008). The hippocampus as a "stupid," domain-specific module: Implications for theories of recent and remote memory, and of imagination. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *62*(1), 62.

Mullally, S. L., & Maguire, E. A. (2014). Memory, imagination, and predicting the future: a common brain mechanism? *The Neuroscientist*, *20*(3), 220–234.

Murphy, T. H., & Corbett, D. (2009). Plasticity during stroke recovery: from synapse to behaviour. *Nature Reviews Neuroscience*, *10*(12), 861.

Murty, V. P., Ritchey, M., Adcock, R. A., & LaBar, K. S. (2010). fMRI studies of successful emotional memory encoding: A quantitative meta-analysis. *Neuropsychologia*, *48*(12), 3459–3469.

N

Nadel, L., & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology*, *7*(2), 217–227.

Nadel, L., Samsonovich, A., Ryan, L., & Moscovitch, M. (2000). Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results. *Hippocampus*, *10*(4), 352–368.

Nairne, J. S., & Pandeirada, J. N. S. (2008). Adaptive memory: Remembering with a stone-age brain. *Current Directions in Psychological Science*, *17*(4), 239–243.

Neisser, U., & Harsch, N. (1992). Phantom flashbulbs: False recollections of hearing the news about Challenger. In *Affect and accuracy in recall: Studies of "flashbulb" memories* (pp. 9–31). Cambridge University Press.

Newman, M. (2018). *Networks*. Oxford university press.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*(23), 8577–8582.

Nir, Y., Staba, R. J., Andrillon, T., Vyazovskiy, V. V, Cirelli, C., Fried, I., & Tononi, G. (2011). Regional slow waves and spindles in human sleep. *Neuron*, *70*(1), 153–169.

Nyberg, L., Habib, R., McIntosh, A. R., & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Academy of Sciences*, *97*(20), 11120–11124.

Nyberg, L., Mclntosh, A. R., Houle, S., Nilsson, L.-G., & Tulving, E. (1996). Activation of medial temporal structures during episodic memory retrieval. *Nature*, *380*(6576), 715–717.

Nyffeler, T., Wurtz, P., Lüscher, H.-R., Hess, C. W., Senn, W., Pflugshaupt, T., … Müri, R. M. (2006). Repetitive TMS over the human oculomotor cortex: comparison of 1-Hz and theta burst stimulation. *Neuroscience Letters*, *409*(1), 57–60.

O O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: psychophysiological interactions and functional connectivity. *Social Cognitive and Affective Neuroscience*, *7*(5), 604–609.

O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2014). Complementary learning systems. *Cognitive Science*, *38*(6), 1229–1248.

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends in Cognitive Sciences*, *6*(12), 505–510.

O'Shea, J., & Walsh, V. (2007). Transcranial magnetic stimulation. *Current Biology*, *17*(6), R196–R199.

Ogawa, S., Lee, T.-M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, *87*(24), 9868–9872.

Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., … Yamadori, A. (2003). Thinking of the future and past: The roles of the frontal pole and the medial temporal lobes. *Neuroimage*, *19*(4), 1369–1380.

Ong, J. L., Lo, J. C., Chee, N. I. Y. N., Santostasi, G., Paller, K. A., Zee, P. C., & Chee, M. W. L. (2016). Effects of phase-locked acoustic stimulation during a nap on EEG spectra and declarative memory consolidation. *Sleep Medicine*, *20*, 88–97.

Öngür, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex*, *10*(3), 206–219.

Oudiette, D., Antony, J. W., Creery, J. D., & Paller, K. A. (2013). The role of memory reactivation during wakefulness and sleep in determining which memories endure. *Journal of Neuroscience*, *33*(15), 6672–6678.

P Palombo, D. J., Hayes, S. M., Peterson, K. M., Keane, M. M., & Verfaellie, M. (2016). Medial temporal lobe contributions to episodic future thinking: Scene construction or future projection? *Cerebral Cortex*, *28*(2), 447–458.

Park, R. J., Goodyer, I. M., & Teasdale, J. D. (2002). Categoric overgeneral autobiographical memory in adolescents with major depressive disorder. *Psychological Medicine*, *32*(2), 267–276.

Pasupathy, A., & Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature*, *433*(7028), 873–876.

Patihis, L., Frenda, S. J., LePort, A. K. R., Petersen, N., Nichols, R. M., Stark, C. E. L., … Loftus, E. F. (2013). False memories in highly superior autobiographical memory individuals. *Proceedings of the National Academy of Sciences*, *110*(52), 20947–20952.

Payne, J. D., Jackson, E. D., Hoscheidt, S., Ryan, L., Jacobs, W. J., & Nadel, L. (2007). Stress administered prior to encoding impairs neutral but enhances emotional long-term episodic memories. *Learning & Memory*, *14*(12), 861–868.

Payne, J. D., Schacter, D. L., Propper, R. E., Huang, L.-W., Wamsley, E. J., Tucker, M. A., … Stickgold, R. (2009). The role of sleep in false memory formation. *Neurobiology of Learning and Memory*, *92*(3), 327–334.

Peigneux, P., Laureys, S., Fuchs, S., Collette, F., Perrin, F., Reggers, J., … Aerts, J. (2004). Are spatial memories strengthened in the human hippocampus during slow wave sleep? *Neuron*, *44*(3), 535–545.

Peigneux, P., Maquet, P., Van der Linden, M., Meulemans, T., Degueldre, C., Delfiore, G., … Franck, G. (1999). Left inferior frontal cortex is involved in probabilistic serial reaction time learning. *Brain and Cognition*, *40*(1), 215–219.

Penfield, W., & Milner, B. (1958). Memory deficit produced by bilateral lesions in the hippocampal zone. *AMA Archives of Neurology & Psychiatry*, *79*(5), 475–497.

Peters, M. J. V, Jelicic, M., Verbeek, H., & Merckelbach, H. (2007). Poor working memory predicts false memories. *European Journal of Cognitive Psychology*, *19*(2), 213–232.

Petersson, K. M., Forkstam, C., & Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cognitive Science*, *28*(3), 383–407.

Peyrache, A., Khamassi, M., Benchenane, K., Wiener, S. I., & Battaglia, F. P. (2009). Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature Neuroscience*, *12*(7), 919–926.

Phan, K. L., Taylor, S. F., Welsh, R. C., Decker, L. R., Noll, D. C., Nichols, T. E., … Liberzon, I. (2003). Activation of the medial prefrontal cortex and extended amygdala by individual ratings of emotional arousal: a fMRI study. *Biological Psychiatry*, *53*(3), 211–215.

Phelps, E. A. (2004). Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology*, *14*(2), 198–202.

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron*, *48*(2), 175–187.

Pierce, B. H., & Kensinger, E. A. (2011). Effects of emotion on associative recognition: valence and retention interval matter. *Emotion*, *11*(1), 139–144.

Poldrack, R. A., Clark, J., Pare-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, *414*(6863), 546–550.

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, *310*(5756), 1963–1966.

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in Cognitive Sciences*, *17*(5), 230–240.

Poppenk, J., & Moscovitch, M. (2011). A Hippocampal Marker of Recollection Memory Ability among Healthy Young Adults: Contributions of Posterior and Anterior Segments. *Neuron*, *72*(6), 931–937.

Poser, B. A., Versluis, M. J., Hoogduin, J. M., & Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine*, *55*(6), 1227–1235.

Power, J. D., Schlaggar, B. L., Lessov-Schlaggar, C. N., & Petersen, S. E. (2013). Evidence for hubs in human functional brain networks. *Neuron*, *79*(4), 798–813.

Preston, A. R., & Eichenbaum, H. (2013). Interplay of Hippocampus and Prefrontal Cortex in Memory. *Current Biology*, *23*(17), R764–R773.

Price, A. R., Bonner, M. F., Peelle, J. E., & Grossman, M. (2015). Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *The Journal of Neuroscience*, *35*(7), 3276–3284.

Q Quirk, G. J., & Beer, J. S. (2006). Prefrontal involvement in the regulation of emotion: convergence of rat and human studies. *Current Opinion in Neurobiology*, *16*(6), 723–727.

R Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676–682.

Ranganath, C., Heller, A., Cohen, M. X., Brozinsky, C. J., & Rissman, J. (2005). Functional connectivity with the hippocampus during successful memory formation. *Hippocampus*, *15*(8), 997–1005.

Rasch, B., Büchel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science*, *315*(5817), 1426–1429.

Redondo, R. L., & Morris, R. G. M. (2011). Making memories last: the synaptic tagging and capture hypothesis. *Nature Reviews Neuroscience*, *12*(1), 17.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *Classical conditioning: Current research and theory*.

Ribot, T. (1882). *Diseases of memory: An essay in the positive psychology* (Vol. 43). D. Appleton.

Rice, G. E., Lambon Ralph, M. A., & Hoffman, P. (2015). The Roles of Left Versus Right Anterior Temporal Lobes in Conceptual Knowledge: An ALE Meta-analysis of 97 Functional Neuroimaging Studies. *Cerebral Cortex* , *25*(11), 4374–4391.

Ridding, M. C., & Rothwell, J. C. (2007). Is there a future for therapeutic use of transcranial magnetic stimulation? *Nature Reviews Neuroscience*, *8*(7), 559.

Rimmele, U., Davachi, L., Petrov, R., Dougal, S., & Phelps, E. A. (2011). Emotion enhances the subjective feeling of remembering, despite lower accuracy for contextual details. *Emotion*, *11*(3), 553–562.

Rimmele, U., Davachi, L., & Phelps, E. A. (2012). Memory for time and place contributes to enhanced confidence in memories for emotional events. *Emotion*, *12*(4), 834–846.

Ripollés, P., Marco-Pallarés, J., Hielscher, U., Mestres-Missé, A., Tempelmann, C., Heinze, H.-J., … Noesselt, T. (2015). The Role of Reward in Word Learning and Its Implications for Language Acquisition. *Current Biology*, *24*(21), 2606–2611.

Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *Neuroimage*, *23*(2), 752–763.

Ritchey, M., Dolcos, F., & Cabeza, R. (2008). Role of amygdala connectivity in the persistence of rmotional memories over time: An event-related fMRI investigation. *Cerebral Cortex*, *18*(11), 2494–2504.

Roberts, R. P., Schacter, D. L., & Addis, D. R. (2017). Scene construction and relational processing: Separable constructs? *Cerebral Cortex*, *28*(5), 1729–1732.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803.

Roozendaal, B., & McGaugh, J. L. (1997). Basolateral amygdala lesions block the memory-enhancing effect of glucocorticoid administration in the dorsal hippocampus of rats. *European Journal of Neuroscience*, *9*(1), 76–83.

Rossi, S., Hallett, M., Rossini, P. M., & Pascual-Leone, A. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology*, *120*(12), 2008–2039.

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, *52*(3), 1059–1069.

Rudoy, J. D., Voss, J. L., Westerberg, C. E., & Paller, K. A. (2009). Strengthening individual memories by reactivating them during sleep. *Science*, *326*(5956), 1079.

Rudy, J. W., & Sutherland, R. J. (1995). Configural association theory and the hippocampal formation: An appraisal and reconfiguration. *Hippocampus*, *5*(5), 375–389.

Rumelhart, D. E., & McClelland, J. L. (1986). Parallel distributed processing: explorations in the microstructure of cognition. volume 1. foundations.

Rushworth, M. F. S., Hadland, K. A., Paus, T., & Sipila, P. K. (2002). Role of the Human Medial Frontal Cortex in Task Switching: A Combined fMRI and TMS Study. *Journal of Neurophysiology*, *87*(5), 2577–2592.

S Sakaki, M., & Mather, M. (2012). How reward and emotional stimuli induce different reactions across the menstrual cycle. *Social and Personality Psychology Compass*, *6*(1), 1–17.

Saleem, K. S., Kondo, H., & Price, J. L. (2008). Complementary circuits connecting the orbital and medial prefrontal networks with the temporal, insular, and opercular cortex in the macaque monkey. *Journal of Comparative Neurology*, *506*(4), 659–693.

Sandrini, M., Cappa, S. F., Rossi, S., Rossini, P. M., & Miniussi, C. (2001). The Role of Prefrontal Cortex in Verbal Episodic Memory: rTMS Evidence. *Journal of Cognitive Neuroscience*, *15*(6), 855-861.

Schabus, M., Gruber, G., Parapatics, S., Sauter, C., Klösch, G., Anderer, P., ... Zeitlhofer, J. (2004). Sleep spindles and their significance for declarative memory consolidation. *Sleep*, *27*(8), 1479–1485.

Schacter, D. L., Guerin, S. A., & Jacques, P. L. S. (2011). Memory distortion: An adaptive perspective. *Trends in Cognitive Sciences*, *15*(10), 467–474.

Schacter, D. L., & Tulving, E. (1994). Memory systems 1994. *Memory Systems 1994*.

Schacter, D. L., & Wagner, A. D. (1999). Medial temporal lobe activations in fMRI and PET studies of episodic encoding and retrieval. *Hippocampus*, *9*(1), 7–24.

Schlichting, M. L., & Preston, A. R. (2016). Hippocampal–medial prefrontal circuit supports memory updating during learning and post-encoding rest. *Neurobiology of Learning and Memory*, *134*, 91–106.

Schmitter, S., Diesch, E., Amann, M., Kroll, A., Moayer, M., & Schad, L. R. (2008). Silent echo-planar imaging for auditory FMRI. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *21*(5), 317–325.

Schnider, A. (2003). Spontaneous confabulation and the adaptation of thought to ongoing reality. *Nature Reviews Neuroscience*, *4*(8), 662–671.

Schouten, D. I., Pereira, S. I. R., Tops, M., & Louzada, F. M. (2017). State of the art on targeted memory reactivation: sleep your way to enhanced cognition. *Sleep Medicine Reviews*, *32*, 123–131.

Schulz, H. (2008). Rethinking Sleep Analysis Comment on the AASM Manual for the Scoring of Sleep and Associated Events. *Journal of Clinical Sleep Medicine*, *4*(2), 99–103.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.

Scimeca, J. M., & Badre, D. (2012). Striatal contributions to declarative memory retrieval. *Neuron*, *75*(3), 380–392.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, *20*(1), 11.

Seger, C. A., & Cincotta, C. M. (2006). Dynamics of Frontal, Striatal, and Hippocampal Systems during Rule Learning. *Cerebral Cortex* , *16*(11), 1546–1555.

Seger, C. A., & Miller, E. K. (2010). Category learning in the brain. *Annual Review of Neuroscience*, *33*, 203–219.

Seger, C. A., Poldrack, R. A., Prabhakaran, V., Zhao, M., Glover, G. H., & Gabrieli, J. D. E. (2000). Hemispheric asymmetries and individual differences in visual concept learning as measured by functional MRI. *Neuropsychologia*, *38*(9), 1316–1324.

Seidenbecher, T., Reymann, K. G., & Balschun, D. (1997). A post-tetanic time window for the reinforcement of long-term potentiation by appetitive and aversive stimuli. *Proceedings of the National Academy of Sciences*, *94*(4), 1494–1499.

Shin, L. M., Rauch, S. L., & Pitman, R. K. (2006). Amygdala, medial prefrontal cortex, and hippocampal function in PTSD. *Annals of the New York Academy of Sciences*, *1071*(1), 67–79. https://doi.org/10.1196/annals.1364.007

Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends in Cognitive Sciences*, *14*(10), 464–472.

Siapas, A. G., & Wilson, M. A. (1998). Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep. *Neuron*, *21*(5), 1123–1128.

Simons, J. S., & Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nature Reviews Neuroscience*, *4*(8), 637.

Sirota, A., Csicsvari, J., Buhl, D., & Buzsáki, G. (2003). Communication between neocortex and hippocampus during sleep in rodents. *Proceedings of the National Academy of Sciences*, *100*(4), 2065–2069.

Skaggs, W. E., & McNaughton, B. L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science*, *271*(5257), 1870–1873.

Smeets, T., Otgaar, H., Raymaekers, L., Peters, M. J. V, & Merckelbach, H. (2012). Survival processing in times of stress. *Psychonomic Bulletin & Review*, *19*(1), 113–118.

Smith, A. C., Frank, L. M., Wirth, S., Yanike, M., Hu, D., Kubota, Y., … Brown, E. N. (2004). Dynamic Analysis of Learning in Behavioral Experiments. *The Journal of Neuroscience* , *24*(2), 447–461.

Smith, A. P. R., Stephan, K. E., Rugg, M. D., & Dolan, R. J. (2006). Task and content modulate amygdala-hippocampal connectivity in emotional retrieval. *Neuron*, *49*(4), 631–638.

Spalding, K. N., Jones, S. H., Duff, M. C., Tranel, D., & Warren, D. E. (2015). Investigating the Neural Correlates of Schemas: Ventromedial Prefrontal Cortex Is Necessary for Normal Schematic Influence on Memory. *The Journal of Neuroscience*, *35*(47), 15746–15751.

Spoormaker, V. I., Schröter, M. S., Gleiser, P. M., Andrade, K. C., Dresler, M., Wehrle, R., … Czisch, M. (2010). Development of a large-scale functional brain network during human non-rapid eye movement sleep. *Journal of Neuroscience*, *30*(34), 11379–11387.

Squire, L. R. (1992). Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychological Review*, *99*(2), 195.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of Learning and Memory*, *82*(3), 171–177.

Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current Opinion in Neurobiology*, *5*(2), 169–177.

Squire, L. R., Cohen, N. J., & Nadel, L. (1984). The medial temporal region and memory consolidation: A new hypothesis. *Memory Consolidation: Psychobiology of Cognition*, 185–210.

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor Perspectives in Biology*, *7*(8), a021766.

Squire, L. R., & Zola-Morgan, S. (1988). Memory: brain systems and behavior. *Trends in Neurosciences*, *11*(4), 170–175.

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*(5026), 1380–1386.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, *27*(3), 494–500.

Stam, C. J., & Reijneveld, J. C. (2007). Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomedical Physics*, *1*(1), 3.

Staresina, B. P., Bergmann, T. O., Bonnefond, M., Van Der Meij, R., Jensen, O., Deuker, L., ... Fell, J. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nature Neuroscience, 18(11), 1679.*

Staresina, B. P., & Davachi, L. (2006). Differential encoding mechanisms for subsequent associative recognition and free recall. *The Journal of Neuroscience*, *26*(36), 9162–9172.

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior*, *16*(4), 413–424.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, *437*(7063), 1272.

Stickgold, R., & Walker, M. P. (2013). Sleep-dependent memory triage: evolving generalization through selective processing. *Nature Neuroscience*, *16*(2), 139.

Suddendorf, T., Addis, D. R., & Corballis, M. C. (2011). Mental time travel and shaping of the human mind. *M. Bar*, 344–354.

T

Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., ... Fernández, G. (2006). Declarative memory consolidation in humans: a prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(3), 756–761.

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science*, *14*(5), 455–461.

Talmi, D., & Moscovitch, M. (2004). Can semantic relatedness explain the enhancement of memory for emotional words? *Memory & Cognition*, *32*(5), 742–751.

Tamminen, J., Ralph, M. A. L., & Lewis, P. A. (2013). The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *Journal of Neuroscience*, *33*(39), 15376–15381.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

Teyler, T. J., & DiScenna, P. (1986). The hippocampal memory indexing theory. *Behavioral Neuroscience*, *100*(2), 147.

Thielen, J.-W., Takashima, A., Rutters, F., Tendolkar, I., & Fernández, G. (2015). Transient relay function of midline thalamic nuclei during long-term memory consolidation in humans. *Learning & Memory*, *22*(10), 527–531.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, *94*(26), 14792–14797.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*(5811), 515–518.

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *The Biological Bulletin*, *215*(3), 216–242.

Tononi, G., & Cirelli, C. (2003). Sleep and synaptic homeostasis: a hypothesis. *Brain Research Bulletin*, *62*(2), 143–150.

Tononi, G., & Cirelli, C. (2006). Sleep function and synaptic homeostasis. *Sleep Medicine Reviews*, *10*(1), 49–62.

Touryan, S. R., Marian, D. E., & Shimamura, A. P. (2007). Effect of negative emotional pictures on associative memory for peripheral information. *Memory*, *15*(2), 154–166.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., … Morris, R. G. M. (2007). Schemas and Memory Consolidation. *Science* , *316*(5821), 76–82.

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., … Morris, R. G. M. (2011). Schema-Dependent Gene Activation and Memory Encoding in Neocortex. *Science* , *333*(6044), 891–895.

Tsujimoto, S., Genovesio, A., & Wise, S. P. (2011). Frontal pole cortex: encoding ends at the end of the endbrain. *Trends in Cognitive Sciences*, *15*(4), 169–176.

Tuckey, M. R., & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied*, *9*(2), 101.

Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory*, *1*, 381–403.

Tulving, E., & Markowitsch, H. J. (1998). Episodic and declarative memory: role of the hippocampus. *Hippocampus*, *8*(3), 198–204.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*(1), 273–289.

U

Uddén, J., & Bahlmann, J. (2012). A rostro-caudal gradient of structured sequence processing in the left inferior frontal gyrus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 2023–2032.

Uddin, L. Q., Supekar, K., Amin, H., Rykhlevskaia, E., Nguyen, D. A., Greicius, M. D., & Menon, V. (2010). Dissociable Connectivity within Human Angular Gyrus and Intraparietal Sulcus: Evidence from Functional and Structural Connectivity. *Cerebral Cortex*, *20*(11), 2636–2646.

V

van Buuren, M., Kroes, M. C. W., Wagner, I. C., Genzel, L., Morris, R. G. M., & Fernández, G. (2014). Initial investigation of the effects of an experimentally learned schema on spatial associative memory in humans. *Journal of Neuroscience*, *34*(50), 16662–16670.

van den Heuvel, M. P., Sporns, O., Collin, G., Scheewe, T., Mandl, R. C. W., Cahn, W., ... Kahn, R. S. (2013). Abnormal rich club organization and functional brain dynamics in schizophrenia. *JAMA Psychiatry*, *70*(8), 783–792.

van der Linden, M., Berkers, R. M. W. J., Morris, R. G. M., & Fernández, G. (2017). Angular gyrus involvement at encoding and retrieval is associated with durable but less specific memories. *Journal of Neuroscience*, *37*(39).

Van Der Werf, Y. D., Jolles, J., Witter, M. P., & Uylings, H. B. M. (2003). Contributions of thalamic nuclei to declarative memory functioning. *Cortex*, *39*(4), 1047–1062.

van Dongen, E. V, Kersten, I. H. P., Wagner, I. C., Morris, R. G. M., & Fernández, G. (2016). Physical exercise performed four hours after learning improves memory retention and increases hippocampal pattern similarity during retrieval. *Current Biology*, *26*(13), 1722–1727.

van Dongen, E. V, Takashima, A., Barth, M., Zapp, J., Schad, L. R., Paller, K. A., & Fernández, G. (2012). Memory stabilization with targeted reactivation during human slow-wave sleep. *Proceedings of the National Academy of Sciences* , *109*(26), 10575–10580.

van Dongen, E. V, Thielen, J.-W., Takashima, A., Barth, M., & Fernández, G. (2012). Sleep supports selective retention of associative memories based on relevance for future utilization. *PloS One*, *7*(8), e43426.

van Kesteren, M. T. R., Beul, S. F., Takashima, A., Henson, R. N., Ruiter, D. J., & Fernández, G. (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: From congruent to incongruent. *Neuropsychologia*, *51*(12), 2352–2359.

van Kesteren, M. T. R., Fernández, G., Norris, D. G., & Hermans, E. J. (2010). Persistent schema-dependent hippocampal-neocortical connectivity during memory encoding and postencoding rest in humans. *Proceedings of the National Academy of Sciences*, *107*(16), 7550–7555.

van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., & Fernández, G. (2010). Retrieval of associative information congruent with prior knowledge is related to increased medial prefrontal activity and connectivity. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(47), 15888–94.

van Kesteren, M. T. R., Rijpkema, M., Ruiter, D. J., Morris, R. G. M., & Fernández, G. (2014). Building on prior knowledge: schema-dependent encoding processes relate to academic performance. *Journal of Cognitive Neuroscience*, *26*(10), 2250–2261.

van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, *35*(4), 211–9.

van Marle, H. J. F., Hermans, E. J., Qin, S., & Fernández, G. (2009). From Specificity to Sensitivity: How Acute Stress Affects Amygdala Processing of Biologically Salient Stimuli. *Biological Psychiatry*, *66*(7), 649–655.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, *277*(5324), 376–380.

Venkatraman, V., Rosati, A. G., Taren, A. A., & Huettel, S. A. (2009). Resolving Response, Decision, and Strategic Control: Evidence for a Functional Topography in Dorsomedial Prefrontal Cortex. *The Journal of Neuroscience*, *29*(42), 13158–13164.

Vertes, R. P., Hoover, W. B., Do Valle, A. C., Sherman, A., & Rodriguez, J. J. (2006). Efferent projections of reuniens and rhomboid nuclei of the thalamus in the rat. *Journal of Comparative Neurology*, *499*(5), 768–796.

Vertes, R. P., Hoover, W. B., Szigeti-Buck, K., & Leranth, C. (2007). Nucleus reuniens of the midline thalamus: link between the medial prefrontal cortex and the hippocampus. *Brain Research Bulletin*, *71*(6), 601–609.

Vilberg, K. L., & Rugg, M. D. (2014). Temporal dissociations within the core recollection network. *Cognitive Neuroscience*, *5*(2), 77–84.

Vossen, P., Maks, I., Segers, R., van der Vliet, H., Moens, M.-F., Hofmann, K., … de Rijke, M. (2013). The Cornetto Database : Architecture and User-Scenarios. In P. Spyns & J. Odijk (Eds.), *Essential Speech and Language Technology for Dutch* (pp. 165–184). Springer Berlin Heidelberg.

W

Wagner, A. D., Paré-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron*, *31*(2), 329–338.

Wagner, I. C., van Buuren, M., Kroes, M. C. W., Gutteling, T. P., van der Linden, M., Morris, R. G., & Fernández, G. (2015). Schematic memory components converge within angular gyrus during retrieval. *eLife*, *4*.

Wagner, U., Hallschmid, M., Rasch, B., & Born, J. (2006). Brief sleep after learning keeps emotional memories alive for years. *Biological Psychiatry*, *60*(7), 788–790.

Waldhauser, G. T., Braun, V., & Hanslmayr, S. (2016). Episodic memory retrieval functionally relies on very rapid reactivation of sensory information. *The Journal of Neuroscience*, *36*(1), 251–260.

Wang, J. X., Rogers, L. M., Gross, E. Z., Ryals, A. J., Dokucu, M. E., Brandstatt, K. L., … Voss, J. L. (2014). Targeted enhancement of cortical-hippocampal brain networks and associative memory. *Science*, *345*(6200), 1054–1057.

Wang, Q., Luo, S., Monterosso, J., Zhang, J., Fang, X., Dong, Q., & Xue, G. (2014). Distributed Value Representation in the Medial Prefrontal Cortex during Intertemporal Choices. *The Journal of Neuroscience* , *34*(22), 7522–7530.

Warren, D. E., Jones, S. H., Duff, M. C., & Tranel, D. (2014). False recall is reduced by damage to the ventromedial prefrontal cortex: implications for understanding the neural correlates of schematic memory. *The Journal of Neuroscience*, *34*(22), 7677–82.

Watkins, E., & Teasdale, J. D. (2001). Rumination and overgeneral memory in depression: effects of self-focus and analytic thinking. *Journal of Abnormal Psychology*, *110*(2), 353–357.

Watkins, P. C., Vache, K., Verney, S. P., & Mathews, A. (1996). Unconscious mood-congruent memory bias in depression. *Journal of Abnormal Psychology*, *105*(1), 34–41.

Werner, N. S., Meindl, T., Engel, R. R., Rosner, R., Riedel, M., Reiser, M., & Fast, K. (2009). Hippocampal function during associative learning in patients with posttraumatic stress disorder. *Journal of Psychiatric Research*, *43*(3), 309–318.

Wilhelm, I., Diekelmann, S., Molzow, I., Ayoub, A., Mölle, M., & Born, J. (2011). Sleep selectively enhances memory expected to be of future relevance. *Journal of Neuroscience*, *31*(5), 1563–1569.

Williams, J. M. G., Barnhofer, T., Crane, C., Herman, D., Raes, F., Watkins, E., & Dalgleish, T. (2007). Autobiographical memory specificity and emotional disorder. *Psychological Bulletin*, *133*(1), 122–148.

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, *265*(5172), 676–679.

Winocur, G., Moscovitch, M., & Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal–neocortical interactions. *Neuropsychologia*, *48*(8), 2339–2356.

X

Xu, S., Jiang, W., Poo, M., & Dan, Y. (2012). Activity recall in a visual cortical ensemble. *Nature Neuroscience*, *15*(3), 449–455.

Xu, W., & Südhof, T. C. (2013). A neural circuit for memory specificity and generalization. *Science*, *339*(6125), 1290–1295.

Y

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525.

Zeithamova, D., Dominick, A. L. L., & Preston, A. R. R. (2012). Hippocampal and Ventral Medial Prefrontal Activation during Retrieval-Mediated Learning Supports Novel Inference. *Neuron*, *75*(1), 168–179.

Z

Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., … Dong, Q. (2010). Individual differences in false memory from misinformation: Cognitive factors. *Memory*, *18*(5), 543–555.

## Supplemental materials and methods to Chapter 3

***Participant instructions***

Below, we provide a detailed overview of the instructions that were given to the participants before a) listening to the word lists and subsequently recalling them, and b) recognizing these words among presented lures. The instructions are English translation of the original Dutch.

Recall instructions given verbally:

---

You will hear lists of words, with this headset [experimenter points to headset]. You can keep the headset on at all times. The words will be mentioned one by one at a relatively high speed. Listen carefully to these words and try to remember them. As soon as the voice has stopped reading the words, you can immediately try to mention as many words as possible from this list out loud. You don't have to pay attention to the order in which the words were mentioned, as long as you recall as many as possible. I will write along on my scoring form, but I will also record what you say with this microphone [experimenter points to microphone] so we can analyze and check later what you said. Therefore, you don't have to pay attention to me. When you think you have mentioned all the words you remember and you don't remember any more words, you can press the button to hear the next list.

---

Recall instructions presented on screen:

---

Welcome in this experiment! You will hear 18 lists of words. They will be read out at a relatively high speed. Try to remember these words. After the complete list has been read, mention as many words from the list as you remember. Mention them even if you are not sure, but try not to guess. Press [ENTER] to start. Good luck!

---

Recognition instructions given verbally:

---

You will see words on the screen. Try to remember for each word whether it is old (have heard it before in the experiment) or new (did not hear it before in the experiment). Also try to assess how certain you are: are you certain the word is old or new, fairly certain the word is old or new, or only maybe the word is old or new. As you see, you have six options, and look carefully at which button belongs to which answer option, and which side – left or right - is old or new. Also try to only use the six buttons that correspond to the six answer options. Good luck!

---

Recognition instructions presented on screen:

---

Welcome in this experiment! You will see words. Please respond to each word whether you have heard it before {old} or whether you did not hear it before {new}. Also assess how certain you are of your answer.
Give your answer in the following manner:
Use the keys [A], [S] en [D] if the word is old.
[A]: “certainly old” [S]: “probably old”,[D]: “maybe old”.
Use the keys [J], [K] en [L] if the word is new.
[J]: “maybe new” [K]: “probably new,[L]: “certainly new”.
You will see 126 words in total.
Press [ENTER] to start. Good luck!

---

***Debriefing questionnaires***

Below, we list the debriefing questions that were presented to the participants after completing the experiment. The instructions are English translation of the original Dutch.

---

**Sleep and alcohol intake**

The next questions will cover your experienced quality of sleep last night. Indicate the extent to which the following statements apply to you, were 1 = not all applicable and 6 = completely applicable.

1. I slept well last night.

1 2 3 4 5 6

2. I am feeling sleepy today.

1 2 3 4 5 6

3. I feel well-rested today.

1 2 3 4 5 6

Provide when answering the following questions an estimate as accurately as possible:

1. Amount of hours I slept last night:

2. Time of going to sleep:

3. Time of waking up:

4. Amount of alcoholic drinks I consumed within the last 24 hours:

---

---

**Task knowledge**

You participated in the so-called DRM-task, where you listen to lists of words, which you then subsequently have to verbally mention and recognize. Think back about when you were doing this task. Did you know already about this task? Could you answer the following question as honestly and accurately as possible?

Did you know, prior or during this task that the word lists were constructed such that it would create a false memory for one word that was left out?

- No
- Yes, prior to performing the task
- Yes, while listening to the word lists
- Yes, while recognizing the words
- I did not do this task

---

## Research data management

This thesis is based on the results of human studies, which were conducted in accordance with the principles of the Declaration of Helsinki. The medical and ethical review board Committee on Research Involving Human Subjects Region Arnhem Nijmegen, Nijmegen, the Netherlands has given approval to conduct these studies

All collected electronic data was backed up on external hard drives, and subsequently archived at the central archive of the Donders Centre for Cognitive Neuroimaging. All paper materials were collected in labeled boxes and stored at the Department of Cognitive Neuroscience, RadboudUMC. Additionally, the electronic data reported in Chapter 4 was stored in an electronic project archive folder with the code: 3013040.40. All data (paper and electronic) from Chapter 3 were entered in the online electronic data management software package called Castor EDC. Data management and monitoring for this study were also performed within Castor EDC. An audit trail was incorporated to provide evidence of the activities that have altered the original data. Data where exported from Castor EDC to comma separated value (.csv) files.

The privacy of the participants was warranted by use of unique individual subject codes. This code correspondended with the codes on paper forms. The correspondence between code and participant identity was always stored and kept separate from any participant data.

The datasets analyzed during these studies are available from the corresponding author on reasonable request.

## Summary of thesis

**[English]**

In this thesis, I investigated how the human brain selectively integrates information from ongoing experience into long-term memory storage, either at initial encoding or during reprocessing in post-encoding offline periods. Specifically, I looked at how this integration of information into memory is modulated at encoding by the emotional valence of information, the relation of information to an established memory schema, or regularities in presented information across repeated learning experiences. Furthermore, memory integration may be modulated at post-encoding offline periods, by the reprocessing of encoded information. In the reported empirical investigations, I measured human behavior in terms of performance accuracy on memory tasks, and related that to brain activity and connectivity as measured with functional Magnetic Resonance Imaging (fMRI; Chapters 2, 4 & 5). Furthermore, I used Transcranial Magnetic Stimulation (TMS) to experimentally perturb brain regions and then measured the behavioral effects of such perturbation (Chapter 3). Here, I briefly summarize the main findings of the experimental studies that form the main body of work in this thesis.

First, I looked at the influence of the emotional valence of associations on the encoding and later retention of these associations. The study, described in **Chapter 2,** aimed to answer the question: **How are inter-individual differences in emotional memory explained by activity and connectivity during encoding?** Here, a large group of healthy male participants were scanned while encoding associations of face-photographs and written identity words that were of a neutral ('driver') or negative ('murderer') valence. Subsequently, memory was tested by prompting participants to retrieve the identity words corresponding to each face. Whereas in both valence categories a similar number of faces were labeled correctly with 'neutral' and 'negative' identities (gist memory), specific associations were found to be less accurately remembered when the identity was negative compared to neutral (specific memory). This pattern of results suggests a reduction in memory specificity for associations containing a component with a negative valence. The encoding of these negative associations was paired with a selective increase in activity in a region called the medial prefrontal cortex (mPFC)

as well as an increase in the connectivity of this region with the hippocampus. Individual differences in valence-specific neural connectivity between mPFC and hippocampus were predictive of this valence-specific reduction of memory specificity. The relationship between loss of emotional memory specificity and medial prefrontal-hippocampal connectivity is in line with the hypothesized role of a medial prefrontal-hippocampal circuit in regulating memory specificity.

Second, I looked at the influence of prior knowledge, or schemas on the encoding and retention of information. Knowledge extracted across previous experiences, or schemas, benefit encoding and retention of such congruent information. However, they can also reduce specificity and augment memory for semantically related, but false information. The mPFC has been ascribed the function of leveraging this prior knowledge to influence encoding and retrieval, based on imaging and patient studies**.** Therefore, in **Chapter 3** I set out to answer the question: **What influence does perturbation of the mPFC have on schema-related memory errors?** I used the Deese-Roediger-McDermott (DRM) paradigm, where the studying of words that fit a common semantic schema are found to induce false memories for words that are congruent with the given schema, but were not actually studied (so-called critical lures). Transcranial magnetic stimulation was used to transiently perturb ongoing mPFC processing immediately before participants performed the DRM-task. We observed the predicted reduction in false recall of the critical lures after mPFC perturbation, compared to two control groups, whereas veridical recall and recognition memory performance remained similar across groups. The data provides initial causal evidence for a role of the mPFC in biasing the assimilation of new memories and their consolidation as a function of prior knowledge.

Third, I looked at how regularities across repeated episodes are integrated into memory systems in the brain. Knowledge is acquired by generalization and integration across repeated learning experiences, which can then be applied to future similar instances. In **Chapter 4**, I set out to answer the question: **What brain regions are involved in the accumulation and updating of novel generalized knowledge structures?** I systematically tracked how a linguistic associative knowledge structure was gradually acquired when learning an abstract artificial language organized by higher-order associative regularities. During learning, we

found that activity in the left inferior frontal gyrus correlated with knowledge updating (as estimated by a learning model) in response to feedback, as well as in response to the extent that accumulated knowledge (also estimated by said learning model) was available during retrieval. In the caudate nucleus, activity correlated with the availability of recently acquired knowledge during retrieval, suggesting it supports the initial retrieval of knowledge. Furthermore, we found that activity in a set of regions, including the mPFC and hippocampus, scaled with accumulated knowledge during feedback presentation, which might be indicative of increased generalization of the hierarchical knowledge structure. Together, these results provide an insight into how knowledge structures are gradually integrated into the neocortex across repeated learning experiences.

Fourth, I looked at the influence of memory reactivation during sleep on brain connectivity and memory retention. Memory reactivation is thought to contribute to the functional reorganization of neural memory traces and memory retention. In **Chapter 5**, I aimed to answer the research question: **What brain connectivity changes are induced by cued memory reactivation during sleep and are they related to memory stabilization?** In this study, participants studied visual object-location associations up to criterion inside the MRI-scanner. The object-location associations were presented together with a characteristic sound. Participants then went to sleep inside the scanner. During periods of deep sleep, marked by slow-waves as observed by polysomnography, half of the learned object-location associations were reactivated by presenting sounds that had been associated with object-location associations. In addition, sounds were presented that had not been previously associated with any object-location association (e.g. control sounds). As a result, I found that the occipital cortex, a visual region, displayed greater network integration (as measured by the participation coefficient, a graph network measure) in response to auditory memory cues versus control sounds. This region in the visual cortex is presumably involved in representing low-level visual aspects of the encoded memories. Moreover, it overlaps with the set of regions found to be active in response to the initial pre-sleep retrieval of the memory trace. This suggests that aspects of the original memory trace are reprocessed in response to the cue, and I showed here that this reprocessing is related to overall memory stabilization. Specifically, the reported increase in network participation of the

occipital cortex was related to the difference in memory performance between the post-sleep and pre-sleep test, indicating that such memory reprocessing might be beneficial for memory retention. Furthermore, the occipital cortex displayed enhanced connectivity with mnemonic regions, namely the hippocampus, parahippocampal gyrus, thalamus and mPFC, during cue sound presentation. Together, these results suggest a neural mechanism whereby cue-induced replay during sleep increases the integration of task-relevant perceptual regions with mnemonic regions. This cross-regional integration may be instrumental for the consolidation and long-term storage of enduring memories.

Our brains have evolved to select information it deems relevant to maintain for future instances. As such, information is not simply copied verbatim into a long-term memory store, but several factors influence the extent to which information is selectively encoded, like the emotional valence of incoming information, its congruency with an established schema, and regularities across episodes. Moreover, following initial encoding, the brain uses similar heuristics to selectively reprocess memories during sleep, providing another window for the brain to selectively integrate information into a long-term memory store. The factors identified in these studies influence the extent to which information is integrated into our memory store, but also the extent to which these memories are processed and represented across either a hippocampal-neocortical network or a neocortical network. Hopefully, the studies reported in this thesis will have shed some light onto how our brains achieve this impressive feat.

**[Nederlands]**

In dit proefschrift heb ik onderzocht hoe het menselijk brein in staat is om van alle informatie waaraan het in het dagelijks leven wordt blootgesteld, selectief informatie in de lange-termijn opslag van het geheugen te integreren. Deze selectieve integratie vindt plaats enerzijds tijdens de initiële blootstelling aan deze informatie (encoderen), anderzijds tijdens periodes die volgen wanneer het brein niet direct meer met de taak bezig is (off-line consolidatie). Ook heb ik onderzocht hoe de integratie van informatie tijdens het encoderen wordt gemoduleerd door de emotionele valentie van informatie, de relatie van bepaalde informatie tot een bestaand geheugenschema, en de consistentie van informatie over herhaalde presentaties binnen een studeerperiode. In de gerapporteerde empirische studies werd het menselijk gedrag gemeten door middel van de accuraatheid van prestaties op geheugenmaten, en deze geheugenprestaties werden vervolgens gerelateerd aan hersenactiviteit en connectiviteit zoals gemeten met functioneel Magnetische Resonantie Imaging (fMRI; Hoofdstukken 2, 4 & 5). Ook heb ik gebruik gemaakt van Transcranieel Magnetische Stimulatie (TMS) om experimenteel de functies van bepaalde hersengebieden te verstoren en de resulterende gedragseffecten te meten (zie Hoofdstuk 3). In deze sectie geef ik een korte samenvatting van de belangrijkste bevindingen van de empirische studies.

Allereerst heb ik gekeken naar de invloed van emotionele valentie van associaties op het encoderen en onthouden van dezelfde associaties. Deze studie, beschreven in **Hoofdstuk 2**, had als doel om de volgende vraag te beantwoorden: **Hoe worden inter-individuele verschillen in emotioneel geheugen verklaard door hersenactiviteit en connectiviteit tijdens het encoderen?** In deze studie heb ik een grote groep gezonde mannelijke proefpersonen gescanned terwijl ze werden blootgesteld aan foto's van gezichten die gepaard gingen met geschreven identiteitswoorden, welke enerzijds neutraal ('chauffeur') en anderzijds negatief ('moordenaar') waren. Vervolgens werd hun geheugen getest door ze te vragen om bij elk gezicht de bijbehorende identiteitswoorden te plaatsen. Waar in beide valentie-categorieën even zoveel gezichten met een identiteitswoord uit dezelfde valentie-categorie werden gecombineerd (gist geheugen), was het geheugen voor het specifieke identiteitswoord minder goed voor de negatieve associaties (specifiek geheugen). Deze resultaten suggereren

dat er een reductie in geheugenspecificiteit plaatsvindt voor associaties die een component bevatten met een negatieve valentie. Het encoderen van deze negatieve associaties ging gepaard met een selectieve toename in activiteit in de mPFC, alsmede connectiviteit tussen mPFC en de hippocampus. Individuele verschillen in neurale connectiviteit tussen de mPFC en de hippocampus tijdens het encoderen van negatieve vergeleken met neutrale associaties waren gerelateerd aan deze reductie in geheugenspecificiteit voor negatieve associaties. Deze relatie tussen een verlies aan geheugen-specificiteit voor negatieve associaties en connectiviteit tussen de mPFC en de hippocampus is consistent met een voorgestelde rol voor een neuraal circuit bestaande uit de mPFC en de hippocampus in het reguleren van geheugen-specificiteit.

Ten tweede, heb ik gekeken naar de invloed van bestaande kennis, of schemas, op het encoderen en onthouden van informatie. Kennis die wordt ge-extraheerd uit voorgaande ervaringen, oftewel schemas, dragen bij aan het encoderen en onthouden van congruente informatie. Echter, bestaande kennis kan ook bijdragen aan een reductie in specificiteit voor herinneringen aan semantisch gerelateerde informatie. Dit kan zelfs bijdragen aan het herinneren van foutieve informatie. De mPFC wordt verondersteld een rol te spelen in het gebruiken van bestaande kennis om nieuwe informatie te encoderen en terug te halen. In **Hoofdstuk 3**, heb ik daarom geprobeerd een antwoord te geven op de vraag: **Welke invloed heeft een verstoring van de mPFC op schema-gerelateerde geheugenfouten?** Hier heb ik gebruik gemaakt van de zogenaamde Deese-Roediger-McDermott (DRM) paradigma, waar het bestuderen van woorden die bij een gemeenschappelijk semantisch schema passen vervolgens resulteert in foutieve herinneringen voor bepaalde woorden die zeer congruent waren met het bestudeerde schema ('critical lures'), maar die niet in werkelijkheid waren bestudeerd. Bij deze studie heb ik TMS gebruikt om tijdelijk processen in de mPFC te verstoren, voordat proefpersonen deze DRM-taak uitvoerden. We observeerden een reductie in de foutieve recollectie van de 'critical lures' volgend op een verstoring van de mPFC, vergeleken met de twee controlegroepen. Daarentegen waren er geen verschillen in correcte recollecties en herkenning tussen de groepen. Deze data geeft ondersteuning voor een rol van de mPFC in het beïnvloeden van het aanmaken en consolideren van nieuwe herinneringen op basis van bestaande kennis.

Ten derde heb ik bekeken hoe regelmatigheden over herhaalde episodes worden opgepikt en geïntegreerd in onze geheugensystemen in het brein. Zulke kennis wordt aangemaakt door middel van generalisatie en integratie over herhaalde leerervaringen, en kan vervolgens worden toegepast op toekomstige voorvallen. In **Hoofdstuk 4**, heb ik daarom geprobeerd de volgende vraag te beantwoorden: **Welke hersengebieden spelen een rol in de accumulatie en bijwerken van nieuw aangeleerde gegeneraliseerde kennisstructuren?** In deze studie volgde ik systematisch hoe een nieuwe linguïstische associatieve kennisstructuur geleidelijk werd aangeleerd door het bestuderen van een abstracte artificiële taal die zich kenmerkt door gegeneraliseerde associatieve regels. Tijdens het leren vonden we dat activiteit in de linker inferiorale frontale gyrus correleerde met het updaten van kennis (geschat door een leermodel) naar aanleiding van feedback, maar ook met de geaccumuleerde kennis die al beschikbaar was bij het ophalen van kennis op basis van een aanwijzing (zoals ook geschat door een leermodel). In een gebied genaamd de caudate nucleus correleerde activiteit met het aanwezig zijn van recentelijk aangemaakte kennis tijdens het ophalen van dezelfde kennis. Verder vonden we ook dat activiteit in een cluster van gebieden, waaronder de mPFC en de hippocampus, correleerde met de accumuleerde kennis tijdens het aanbieden van feedback, wat zou kunnen samenhangen met een grotere mate van generalisatie van de hierarchische kennisstructuur. Samen geven deze resultaten een inzicht in hoe kennisstructuren geleidelijk worden geïntegreerd in de neocortex over herhaalde leerervaringen.

Als vierde, en laatste, onderzocht ik de invloed van het reactiveren van herinneringen gedurende slaap op hersen-connectiviteit en het latere behoud van deze herinneringen. Het heractiveren van herinneringen tijdens slaap wordt gedacht bij te dragen aan de functionele re-organisatie van neurale geheugensporen en het behoud van geheugen. In **Hoofdstuk 5**, trachtte ik daarom de volgende onderzoeksvraag te beantwoorden: **Wat voor veranderingen in brein-connectiviteit worden er veroorzaakt door geheugen-reactivatie tijdens slaap, en hoe zijn deze veranderingen gerelateerd aan geheugen-stabilisatie?** In deze studie werden proefpersonen gevraagd visuele associaties tussen objecten en locaties te leren in de MRI-scanner. De object-locatie associaties werden gepresenteerd gezamenlijk met een karakteristiek geluid. Vervolgens

gingen de proefpersonen slapen in de scanner. Tijdens periodes van diepe slaap, zoals te zien in karakteristieke grote en langzame hersengolven gemeten met polysomnography, werd de helft van de geleerde object-locatie associaties gereactiveerd door karakteristieke geluiden aan de proefpersonen aan te bieden. Daarnaast werden geluiden gepresenteerd die voorheen niet samen met object-locatie associaties waren aangeboden (zogenaamde controle-geluiden). Mijn bevindingen waren dat de occipitale cortex, een gebied gespecialiseerd in de visuele waarneming, zich kenmerkte door een grotere netwerk-integratie (gemeten met de participatie coëfficient, een graph-theoretische maat) in reactie op de presentatie van karakteristieke geluiden in vergelijking met de controle-geluiden. Dit gebied in de visuele cortex is waarschijnlijk belangrijk in de representatie van de visuele aspecten van de opgeslagen herinneringen. Ook overlapte dit gebied met de gebieden die actief waren in reactie op het eerder ophalen van de herinneringen voor het slapengaan. Dit suggereert dat aspecten van het originele geheugenspoor weer worden herverwerkt in reactie op het geluid. Verder vond ik ook dat deze hernieuwde verwerking is gerelateerd aan de algemene stabilisatie van herinneringen. Namelijk, de gerapporteerde toename in netwerk-integratie van de occipitale cortex was gerelateerd aan het verschil in de accuraatheid van het geheugen tussen testen die enerzijds werden afgenomen voor het slapen gaan en anderzijds afgenomen na het slapen gaan. Dit geeft aan dat het herverwerken van de geheugensporen bijdraagt aan het algehele geheugenbehoud. Bovendien vertoonde de occipitale cortex een toegenomen connectiviteit met geheugengebieden, specifiek de hippocampus, de parahippocampale gyrus, de thalamus en de mPFC tijdens de presentatie van geluiden. Samen, geven deze resultaten een beeld van een neuraal mechanisme waarmee de reactivatie van herinneringen tijdens slaap de integratie van taak-relevante perceptuele gebieden en geheugen-gebieden bevordert. Deze inter-regionale integratie kan instrumentaal zijn voor de consolidatie en opslag van blijvende herinneringen.

Onze hersenen hebben zich ge-evolueerd om selectief informatie te integreren die van belang is om te bewaren voor toekomstige voorvallen. Zodoende wordt informatie niet simpelweg letterlijk gekopieerd naar een lange termijn opslag, maar zijn er verschillende factoren die beïnvloeden hoe en welke informatie selectief wordt opgeslagen. Deze factoren zijn onder meer de emotionele valentie

van informatie, in hoeverre deze informatie consistent is met bestaande kennis of schemas, en regelmatigheden tussen opeenvolgende presentaties. Volgend op het initiële aanleren van informatie, maakt het brein van dezelfde regels gebruik om selectief bepaalde geheugensporen off-line te her-verwerken, een andere mogelijkheid voor het brein om selectief informatie te integreren in een stabiele lange termijn geheugenopslag. De geïdentificeerde factoren in deze studies beïnvloeden de wijze waarop informatie word geïntegreerd in ons geheugenopslag, maar ook in hoeverre deze geheugens worden verwerkt en zijn weerslag vinden in een hippocampaal-neocorticaal netwerk, of een neocorticaal netwerk. Hopelijk dragen de studies in dit proefschrift bij aan het begrip van hoe het brein deze taak volbrengt.

## List of publications

**Journal Publications:**

**Berkers, R.M.W.J.*,** Ekman, M.*, van Dongen, E.V., Takashima, A., Barth, M., Paller, K., Fernández, G. (2018). Cued reactivation during slow-wave sleep induces connectivity changes related to memory stabilization. *Scientific Reports. 8*(1), 16958. *Contributed equally to the manuscript.

Benoit, R.G., **Berkers, R.M.W.J.**, Paulus, P.C. (2018). An adaptive function of mental time travel: Motivating farsighted decisions. *Behavioral & Brain Sciences, 41.*

Van der Linden, M., **Berkers, R.M.W.J.,** Morris, R.G.M. & Fernández, G. (2017). Angular gyrus involvement at encoding and retrieval is associated with durable but less specific memories. *Journal of Neuroscience, 37*(39), 9474-9485.

**Berkers, R.M.W.J.,** van der Linden, M., de Almeida, R.F., Müller, N.C.J., Bovy, L., Dresler, M., Morris, R.G.M., Fernandez, G. (2017). Transient medial prefrontal perturbation reduces false memory formation. *Cortex, 88,* 42 – 52.

**Berkers, R.M.W.J.**, Klumpers, F. & Fernandez, G. (2016). Medial prefrontal-hippocampal connectivity during emotional memory encoding predicts individual differences in the loss of associative memory specificity. *Neurobiology of Learning and Memory, 134*: 44 – 54.

**Berkers, R.M.W.J.**, van Kesteren, M.T.R. (2013). Autobiographical memory transformation across consolidation. *Journal of Neuroscience, 33*(13), 5435-5436.

Van Heugten, C.M., Janssen, E.P.J., Visscher, A.J.M., Wolters Gregorio, G., Smeets, S., **Berkers, R.M.W.J.**, Ponds, R.W.M. (2013). Klinische patienten met niet-aangeboren hersenletsel in de GGZ; inventarisatie van zorgbehoeftes en ontvangen zorg. *Tijdschrift voor Psychiatrie, 55*(9): 665-675.

**Book chapters**

**Berkers, R.M.W.J.,** van Goethem, N., Rutten, K., Blokland, A., Prickaerts, J. (2014). The Medial Temporal Lobe: Toward a Unifying Neuropsychobiological Framework of Recognition and Recall, In: Episodic Memory: Formation, Clinical Disorders and Role of Aging, Nova Science Publishers Inc.

**Reports**

van Heugten, C., Ponds, R., **Berkers, R.M.W.J.**., Smeets, S., Wolters Gregório-Claessens, G. (2012). Zorgmonitor NAH in de GGZ: een onderzoek naar de zorgbehoeftes en zorgverlening voor klinische patiënten met niet aangeboren hersenletsel in de GGZ. Maastricht: Maastricht University.

**Journal articles in revision**

**Berkers, R.M.W.J.,** van der Linden, M., Neville, D.A., van Kesteren, M., Morris, R.G., Murre, J. & Fernández, G. (2019). Neural dynamics of accumulating and updating linguistic knowledge structures. *bioRxiv,* 495168.

Bovy, L., **Berkers, R.M.W.J.**, Pottkämper, J., Varatheesvaran, R., Fernández, G., Tendolkar, I., & Dresler, M. (2019). The influence of transcranial magnetic stimulation of the medial prefrontal cortex on emotional memory schemas. *bioRxiv*, 656348.

Mueller, N., Kohn, N., van Buuren, M., Klijn, N., Emmen, H., **Berkers, R.M.W.J.**, … & Fernandez, G. (2019). Differences in strategic abilities but not associative processes explain memory development. *bioRxiv,* 693895.

Landy, J. F., Jia, M., Ding I. L., Viganola, D. Tierney, W., …, **Berkers, R.M.W.J.**, …., Uhlmann, E. L. (in revision). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results.

**Science communication**

Blog posts for Donders Wonders (www.donderswonders.com):

- "Het Human Brain Project: een tussenstand" (EN/NL), 14.07.2016
- "Computer verslaat mens in spel 'Go'" (EN/NL), 22.02.2016
- "Snel en efficiënt woordjes stampen" (EN/NL) (with Marlieke van Kesteren), 01-10.2015
- "Ik zie, ik zie wat jij niet ziet", 27.11.2015
- "Nut en noodzaak van proefdieren in de neurowetenschappen" (with Lieneke Janssen), 27.7.2015
- "Intelligente machines, bovenmenselijk goed?", 11.05.2015
- "ALS: na de hype" (with Jeanette Mostert), 16.10.2014
- "Voetballen met een hersenschudding", 11.08.2014
- "Marshmallows en WK-koorts", 04.07.2014

# Academic portfolio DGCN

## Academic courses

- Wetenschapsjournalistiek (5 ECTS), 2016
- Mindfulness-Based Stress Reduction voor promovendi (5 ECTS), 2016
- Advanced Mathematics (3 ECTS), 2013-2014
- Basic Mathematics (3 ECTS), 2013
- Neuroimaging II (6 ECTS), 2014
- Neuroimaging I (6 ECTS), 2013
- Toolkit for Transcranial Brain Stimulation (1 ECTS), fall 2013
- Toolkit of Cognitive Neuroscience: Advanced course in fMRI data analysis (1 ECTS), 2012

## Conference presentations

**Berkers, R.M.W.J.***, Ekman, M.*, van Dongen, EV, Takashima, A, Paller, K, Fernández, G. (2016). Cued reactivation in slow-wave sleep induces connectivity changes related to memory stabilization. Oral presentation at International Conference on Memory (ICOM), Budapest, July 2016.

**Berkers, R.M.W.J.,** van der Linden, M., Neville, D., van Kesteren, M., Morris, R., Murre, J. & Fernández, G. (2015). Neural dynamics of updating and accumulating conceptual knowledge. Poster presented at Society for Neuroscience 2015, Chicago, October 2015.

**Berkers, R.M.W.J.**, van der Linden, M., Neville, D., van Kesteren, M., Morris, R., Murre, J. & Fernández, G. (2015). Neural dynamics of updating and accumulating conceptual knowledge. Poster presented at Society for Neuroscience 2015, Chicago, October 2015.

**Berkers, R.M.W.J.,** Klumpers, F., van Wingen, G. & Fernández, G. (2013). Medial Prefrontal Cortex activity is related to individual differences in the encoding of emotional face-occupation associations. Poster presented at Society for Neuroscience 2013, San Diego, November 2013.

**Invited presentations**

**Berkers, R.M.W.J.** Targeted memory reactivation during slow-wave sleep, functional brain connectivity changes and their relation to memory stabilization., The Royal Society. Symposium "The offline brain: understanding memory consolidation and reconsolidation." . Newport Pagnell, UK, January 2017.

**Berkers, R.M.W.J.** "Modulatory influences on neural learning systems and long-term memory formation", Max Planck Institute for Cognition & Brain Sciences. Leipzig, Germany, November 2016.

**Berkers, R.M.W.J. "**Modulatory influences on knowledge acquisition and consolidation.", Department of Psychology, University of Toronto. Toronto, January 2016.

**Berkers, R.M.W.J.** Neural dynamics of linguistic conceptual knowledge accumulation and updating. Symposium "Knowledge representation and concept learning.", Nederlandse Vereniging voor Psychonomie winter conference 2015. Egmond aan Zee, NL, December 2015.

**Berkers, R.M.W.J.** How Schemas and Sleep Modulate Memory Consolidation. Symposium "Modulation of Memory Consolidation", Donders Discussions 2015, Nijmegen, NL, October 2015.

**Berkers, R.M.W.J.** Tracking the neural dynamics of the formation of a novel language schema. Theme 1-meeting Language & Communication. Seminar on 'New Word Learning'. Donders Institute of Brain, Cognition & Behaviour, Nijmegen, NL, June 2014.

**Berkers, R.M.W.J.** Are we stuck in Time? Recalling the past and simulating the future. Weekly centre-wide 'Brown Bag Meetings' at the Donders Centre for Cognitive Neuroimaging, Nijmegen, NL, April 2014.

**Extra-curricular training & certification**

BROK-certification. Basiscursus Regelgeving en Organisatie voor Klinische Onderzoekers

Certified user TMS-laboratory Donders Centre for Cognitive Neuroimaging

Certified user Skyra 3T MRI-scanner

**Teaching & mentoring experience**

***Supervisor of bachelor internships:***
Lonneke van der Kolk, Rafael Fernandez de Almeida, Femke Lammertink.

***Supervisor of lab rotation students:***
Franziska Michels, Grazia di Pisa.

***Co-supervisor of master internship:***
Neda Rashidi-Ranjbar.

***Course coordinator:***
"On papers and publishing – The practice of writing a good academic paper (and getting it published)". A course for PhD students and post-docs at the DGCN (Donders Institute for Brain, Cognition & Behaviour) and IMPRS (Max Planck Institute for Psycholinguistics), Nijmegen, NL.

***Lecturer:***
**"**Toolkit of Cognitive Neuroscience: Transcranial Brain Stimulation", toolkit at the Donders Centre for Cognitive Neuroimaging, Nijmegen, NL.

"Cognitive Neuroimaging": course at the Radboud University for the Biology and Molecular Life Sciences bachelor programs, Nijmegen, NL.

"Brain and learning": high school (VWO) course on memory and the brain (Dutch title: "het geheugen en het brein". "Brain and learning": high school (VWO) course on memory and the brain (Dutch title: "het geheugen en het brein", Nijmegen, NL.

**Societies & committees**

PhD-representative at the Donders Centre for Cognitive Neuroimaging (DCCN)

Member of the Donders PhD-council, representing PhD students in the Donders Graduate School for Cognitive Neuroscience (DGCN) at the Donders Institute (DI).

Chairman and member of the representative advisory employee council (OC) of the Donders Centre for Cognitive Neuroimaging of the Donders Institute.

**Grants & awards**

Memrise Prize. Applied science prize for developing the most efficient method of word learning. ($ 10,000, co-recipient), 2017

Radboud Internationalization Fund Travel Grant for participation in Society for Neuroscience 2015 (€ 700)

Radboud Internationalization Fund Travel Grant for participation in Society for Neuroscience 2013 (€ 718)

**Public media / outreach**

Regular contributor to the Donders Institute blog: "Donders Wonders", writing short articles about brain, cognition and behavior and research for a general audience.

Giving introductory talks on cognitive neuroimaging to high school and bachelor students and organizing tours of the Neuroimaging facilities at the Donders Centre for Cognitive Neuroimaging.

**Other academic activities**

Organizer and moderator of symposium "Modulation of Memory Consolidation" at Donders Discussions 2015, Nijmegen, October 2015.

Co-founder of the "Donders Knowledge Forum", a platform for sharing and preserving knowledge within the institute.

CMO-ethics assistance at the Donders Centre for Cognitive Neuroimaging.

## Curriculum vitae

©MPI CBS

Ruud Martinus Wilhelmus Johannes Berkers was born on 16th of June in 1987 in Venray and grew up in the small agricultural village of Ysselsteyn, in Limburg. There, he completed the VWO high school diploma graduating in both the *Science & Health* and the *Economy & Society* profile. With an interest in biology and human behavior, he enrolled at University College Maastricht to pursue a *Liberal Arts & Sciences* bachelor degree. During his studies, he discovered the possibilities to study the human brain and thinking in a quantitative and scientific manner. In 2007, he studied abroad for a semester at the University of Queensland in Brisbane, Australia. After returning, he graduated cum laude, with a concentration in both life sciences (biology, physiology) and social sciences (psychology).

In the year following the end of his bachelor studies, Ruud worked as a teaching assistant at the University College Maastricht, as a freelance writer for the university newspaper *Observant*, and as a waiter in a restaurant. He then enrolled in the research master program *Cognitive & Clinical Neurosciences*, with a focus on Neuropsychology, at the University of Maastricht to pursue bothresearch and clinical training. Alongside his studies, he worked as a teaching assistant at the Psychology faculty and as a research assistant for a nation-wide study into mental health care for patients with acquired brain damage. Moreover, he obtained the *Top 3 % Award* as one of the top students of his program. As part of his studies, he completed a clinical internship at the Dr. Leo Kannerhuis, a mental health institute that specializes in autism spectrum conditions, gaining a certification in psychodiagnostics. Furthermore, he completed a second research internship in the group of Nikolaus Kriegeskorte at the MRC Cognition & Brain Sciences Unit, using fMRI to study visual object representations in autism spectrum conditions.

After graduation, Ruud joined the *Memory & Emotion* group of professor Guillén Fernández at the Donders Institute for Brain, Cognition & Behaviour at the RadboudUMC in Nijmegen, to pursue his doctoral research studies, presented in this thesis. In addition to research, he also enthusiastically contributed to the academic environment by giving talks to visiting high school students, by popular science throughblogging for *Donders Wonders*, by acting as a PhD-representative and member of the PhD-council as well as as the chairman of the *Onderdeel Commissie* (an employee representative organ), by teaching and coordinating a course on writing papers and publishing for graduate students and by supervising students. Together with a group of researchers from Radboud University & RadboudUMC, he, furthermore, collaborated on a project to design the most effective learning method for acquiring foreign vocabulary, thereby winning the *Memrise Prize*, an international applied science prize, in 2017.

Following his time in Nijmegen, Ruud moved to the Max Planck Institute for Human Cognitive & Brain Sciences in Leipzig to perform postdoctoral research in the *Adaptive Memory* research group of Dr. Roland Benoit on episodic memory, episodic future simulations & decision-making. Here, he successfully ran studies combining behavioural paradigms with psychophysics, fMRI and combined TMS-fMRI, and supervised several bachelor and master students. In order to continue and expand his interest and knowledge in both research and clinical training, Ruud will pursue a career as a clinical psychologist in the near future.

## Acknowledgements

There were many people I have encountered along my PhD-journey that were important in each their own unique respect. To each of you reading this, thank you!

In particular, first of all, I would like to thank my promotor **Guillén** for offering me the opportunity to pursue a PhD project, and for giving me the freedom to explore different research avenues and charter my own path. Your pragmatism, goal-directedness, insight, and drive, helped me make it through, and grow as a person in the process.

Mijn co-promotor **Marieke**, bedankt voor je steun, voor je nuchterheid, voor het luisterend oor, praktische adviezen en aanmoedigingen.

Thanks to all my **collaborators**, without whom the various projects would not have been possible. **Marlieke**, mijn Donders-avontuur begon eigenlijk toen je me in een pub in Cambridge, met een ale, vertelde over schemas en hoe tof het Donders wel niet was. **Matthias**, thanks for working together and your peculiar sense of humor. **Floris**, bedankt voor de samenwerking en het uitstapje naar de 'emos'! **Eelco**, dank dat je me met de slaapdata aan de slag liet gaan, maar vooral dat ik door jou elke donderdag lekker met een bal kon uitrazen. **David**, maestro, thanks for tackling the most complex models and task designs, and brainstorming to come up with the most revolutionary new experiments. **Martin**, thanks for hosting wine & cheese nights. **Nils M.**, thanks for pooling our interests in development and TMS into two fruitful collaborations, and sharing your enthusiasm and energy (whether figuratively, or literally in the shape of your dubious 'energy balls' ;)). I was happy to share my hard-gained TMS-knowledge with **Leonore** in the lab. **Richard M.**, thanks for an inspiring collaboration, transcending species in a quest to discover schemas, together with **Lisa G., Tomonori & Adrian,** be it in Edinburgh, Nijmegen or Chicago. **Jaap**, dank voor de introductie in de wereld van de leermodellen. I would like to thank my brilliant students, who I supervised for a master thesis (**Neda**), bachelor thesis (**Femke, Lonneke**), internship **(Rafael)** and lab rotations (**Grazia, Franziska).** Without the TMS-community (**Lennart, Til, Jim, Thomas, Ian C., Dilene, Dick, Moniek, & Marius** and others) I would not have been able

to explore and venture into brain stimulation territory, I would not have known where to start.

Thanks to the other members of the **Memory & Emotion** group (the 'memos' and the 'emos'), and the warm atmosphere across lunches, labmeetings, pizza-meetings and retreats. In addition to those I already mentioned, I would also like to particularly mention **Isabella, Linda, Susanne, Noortje, Mariët, Marloes, Erno, Nils K., Boris, Daphne E., Daphne T., Christina, Tim v T., Sabine K., Geeralien, Atsuko, Iske, Gabi** and also thank all other students and visitors!

I was involved in several 'side-projects'. I learned a lot from being in the OC with **Sander, Kim, Nils M., & Tineke**. **Mirjam,** samen waren we phd-representatives en daardoor had ik extra veel 'om te gieren!' When **Marlieke** first proposed to start a team (with **Gesa, Anke Marit, Boris, Paul K., Nils M.**) to compete in a world-wide competition to develop the most efficient word-learning method, I never grasped how epic the project would turn out to be. Met **Anke Marit** slaagde ik erin professioneel een vak te coördineren. Lieneke, bedankt dat je me enthousiast aan het bloggen kreeg, en een eigen categorie gaf als 'vaste gastblogger'.

I would also like to thank the non-scientific staff, without whom none of our research would have been possible. **Tildie,** het kloppende hart van het DCCN. **Sandra, Ayse & Nicole**, voor altijd een vriendelijk gezicht bij binnenkomst. **Paul G.**, omdat je me altijd voor de gek hield bij het scannen. **Lucia**, voor de praatjes tussen de bedrijven door. **Arthur** & **Berend**, voor het draaiende houden van een geoliede machine. **Marek, Mike & Uriël**, voor het troubleshooten van mijn technische problemen.

I would like to thank all my **participants** for trying hard to learn impossible languages and for enduring my brain stimulation experiments. Furthermore, I would like to thank my **reading committee** for critically reading and evaluating my thesis manuscript.

Of course, there was also an important social component to my time in Nijmegen and at the Donders. Dank aan mijn kamergenootjes in de eerste jaren in onze

'corner office' op de eerste verdieping (**Marieke S, Arjen, Tim, Maarten & Pim**) voor de nodige en prettige afleiding, and later on, in other offices with **Nietzche, Loek, Payam, Erik, Anke & Hannah.** You all made me happy to be in the office.

The **Donders runners** kept me in shape by running, whether it be after work or running the Bata together with **Sybrine, Sasha, Egbert, Eelco, Verena, Corina, Tim, Anke Marit, Thomas M., Loes, Steffi, Ingar, Maaike, Ivar, Larry, Dan, Fenny, Dan, Frank & Galina. Jeanette**, dank dat je een verdwaalde Limbo op sleeptouw nam om samen door het hele land rond te rennen, en uiteindelijk na de marathon te triomferen in het Olympisch stadion. Ook de andere sportieve activiteiten, zoals blaren op de voeten lopen bij de Vierdaagse met onder anderen **Anke Marit, Mirjam, Jeanette, Tjerk, Larry, Michiel, Desiree, Alina,** leren skieën met **Sybrine, Anke Marit, Sasha, Tjerk, Tim, René, Lisa, Ivar & Roemer,** voetballen op de donderdagavond, het hoogtepunt van mijn week, met onder andere **Eelco, Sasha, Arjen, Egbert, René, Wessel, Pim, Maarten, Christiano Alan, Rasim, Peter K. & Peter K., Monja, Jacob, Izabela,** volleyballen, in de zaal of op het zand, bedankt **Loes, Steffi, Sanne, Larry, Sasha, Linda, Fenny, Dan**, en squash met **Jim, Michelle, Eelco, Verena, Egbert & Tineke G..**

For the **Social Donderians,** the FAD (Friday Afternoon Drink) was a big tradition. There are simply too many people to mention here, but I would like to specifically mention **Sean** whose poems lured me to the Cultuurcafé every friday. Weekly pubquizzes with 'Rank Deficient', including **Ruben, Flora, Sean, Thomas, Richard K., & René**. Met mijn homies **Miriam** en **Linda** hadden we aan de Oranjesingel een Donders party place, met samen koken, yoga in de tuin en de traditionele barbecue tijdens de Vierdaagse. **Miriam**, bedankt voor  je inzichten en peptalks. **Thomas W.**, your peptalks have been important to me too – thank you for your positivity. **Egbert, Adelina, Jelmer & Coco,** for being the first people I visit when back in Nijmegen. **Sasha & Harold,** voor jullie levenswijsheden. **Jeroen, Eelke S., Eelco & Sander,** de sterren van de karaoke. **Lindi & Isi**, the three of us shared many experiences over the years, it is hard to summarize. Koffietjes at the Donders, dinners, gigs and festivals, roadtripping, movie nights – thanks for everything. **Isi,** we started our PhD almost at the same time on the same project, thanks for being on this journey together and being there as my paranymph in the end.

I would also like to thank my other paranymps**. Ari,** eindelijk kom je weer deze kant op en kunnen we weer over schemas, conceptual alignment en brain-to-brain coupling (!) sparren, kan ik je grappen over Limbos weer aanhoren en je kanonskogels in de zaal proberen te ontwijken. **Dan,** thanks for all the adventures, whether it be letting our hair down in Nijmegen, or biking, kayakking and running across Ireland (still bugs me that we did not make it to the top of that mountain ;)).

A big thank you to my **international crew**. Pretty amazing that more than 12 years after first meeting, we are still hanging out. All we need is a location, anywhere in the world, and we reunite and make it ours. **Jenny, Johannes & Julius**, **Janice, Tobi & Clarita, Brenton, Amy & Ayden, Daniel, Ana & Renata, Kayla & Adam, Priscilla & Falco, Mike & Janelle, Kenisha, Guillermo & little Dominik, Ryan & Sarah, Tom & Patrik,** thanks for showing me the world. May the BOA forever be liquid and solvent.

I want to thank my friends from my time in Maastricht. **Ian P.**, thanks for your humor and depth, and discussing religion, philosophy & politics with equal ease and humor as football and movies. My other friends from Maastricht **Miroslav, Federico, Mimi, Karin,** always great to catch up every now and then. Hope I see you again soon.

**Roland,** thanks for taking me on as a postdoc in your lab, years after first meeting in Cambridge. It is great to be a part of a small, independent research group with a talented, positive and diverse group of people, such as **Ann Kristin, Philipp P., Davide, Aroma, Hanna,** and my students **Paula, Marie, Lukas, Philipp K. & Alexandra.**

My friends in Leipzig allowed the tradition of FAD to continue under the guise of 'Biercóles' and 'Biernes', thanks **Alejandro, Nestor, Clara E., Clara K., Pauline, Louise, Adelina, Corrina, Christa, Cornelius, Lena, Alex T., Seyma, Rudi, Ahmad, Glad, Memo** and many others. **Lieneke,** doordat je me gevolgd bent naar Leipzig is het hier gezellig (met een zachte 'G').

Mijn maten thuis in Limburg, **René, Paul, Richard, Willem, Frank, Rogier, Marcel, Bart, Mark**, voor de afwisseling was er altijd jullie platte humor. Doet de boel weer een beetje relativeren ;). We moeten weer vaker een biertje drinken wanneer ik weer in de buurt woon.

Door mijn familie weet ik, waar ook ter wereld, altijd dat er een thuis is. **Rob, Judith & Isabel, Julia & Kiki; Thijs, Maaike & Viggo & Mads; Hans; Daan, Ellen (**en binnenkort jullie kleine meid), bedankt dat jullie er zijn. Mijn ouders **Nettie** en **Jan**, ondanks dat jullie niet altijd helemaal begrepen waar ik nou mee bezig was, of wanneer ik nou klaar was met studeren, zijn jullie er altijd voor me en steunen jullie mij door dik en dun.

**Martina**, you came into my life and changed my world. You gave me a new home, together with your son **Anton** and the cat **Gizmo.** You gave me focus and priority, and are my most critical reviewer and ardent supporter. Excited to see where life takes us next.

## Donders graduate school for cognitive neuroscience

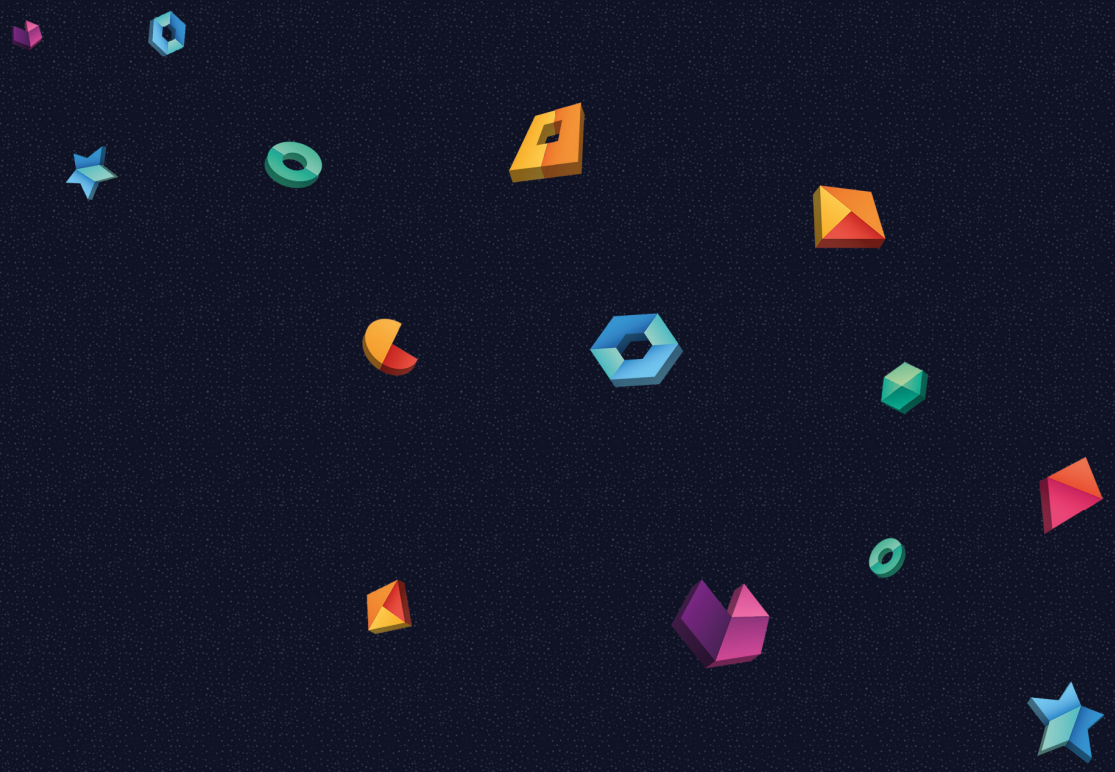
For a successful research institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc.. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology. Specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy. Positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors.

Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy. For more information on the DGCN as well as past and upcoming defenses please visit:

http://www.ru.nl/donders/graduate-school/phd/



DONDERS INSTITUTE

Max Planck Institute for Psycholinguistics

Radboud University

Radboudumc